Quality-aware Resource Management inWireless Networks

Kandaraj Piamrat

Nº d'ordre : 4163 ANNÉE 2010

THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne

pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : Informatique

Ecole doctorale MATISSE

présentée par

Kandaraj PIAMRAT

préparée à l'unité de recherche nº 6074
IRISA / INRIA - Rennes

Quality-aware Resource Management in Wireless Networks

Gestion de ressources basée sur la qualité dans les réseaux sans-fil

Thèse soutenue à Rennes
le 27 septembre 2010

devant le jury composé de :

André-Luc BEYLOT/rapporteur
Professeur à l'INPT/ENSEEIHT

Pascal LORENZ/rapporteur
Professeur à l'Université de Haute Alsace

Thierry TURLETTI/examinateur
Directeur de Recherche à l'INRIA Sophia Antipolis

César VIHO/directeur de thèse
Professeur à l'Université de Rennes 1

Adlen KSENTINI/co-directeur
Maitre de Conférence à l'Université de Rennes 1

Jean-Marie BONNIN/co-directeur
Professeur à Telecom Bretagne

Life is growth. If we stop growing, technically and spiritually, we are as good as dead.
Morihei Ueshiba, Founder of Aikido

Remerciements

C'est avec grand plaisir que je tiens à exprimer ma profonde gratitude à toutes les personnes qui ont fait que cette thèse soit possible.

Tout d'abord, je voudrais remercier les membres du jury d'avoir accepté d'évaluer mes travaux de thèse. Je remercie en particulier M. André-Luc Beylot et M. Pascal Lorenz d'avoir eu la gentillesse d'être rapporteurs et M. Thierry Turletti d'avoir bien voulu juger mon travail.

Je dédie également mes remerciements à tous les membres de l'équipe DIONYSOS, qui m'ont très bien accueilli depuis 3 ans, en particulier à M. Gerardo Rubino pour tous les conseils concernant PSQA et à Kamal pour ceux concernant la technologie 3G. Je remercie Majd qui a été une superbe voisine pendant ces années ainsi que Laura et Kamal pour l'aide apportée dans l'amélioration de ma présentation pour la soutenance. En outre, je tiens à remercier notre assistante d'équipe, Fabienne Cuyolla, pour tous ses soutiens. Je remercie mes amis et collègues à l'IRISA qui ont créé une ambiance conviviale de travail. Certains d'entre eux sont Alexandra, Anthony, Bruno, Nanxing, Raymond, Samira, Sébastien, Sofiane, Tran, Wael. Je tiens à remercier François pour la relecture du résumé en français et pour son soutien tout au long de la thèse.

Un grand remerciement à M. César Viho, mon directeur de thèse, pour son support pendant ces années. Il a su me guider pour mener ce travail à bon terme. Je remercie M. Adlen Ksentini et M. Jean-Marie Bonnin, les discussions techniques avec eux ont grandement contribué à l'avancement et à l'aboutissement de cette thèse.

Enfin et surtout, je me sens redevable envers mes parents, Rumphai et Prakij Piamrat qui m'ont continuellement encouragée et sont toujours là pour moi. Ils ont été un appui indispensable tout au long de mes années d'études. Sans leur soutien continu, cette thèse n'aurait pas été possible. Je leur dédie ce travail.

Contents

Résumé en Français

1. Introduction

Ce chapitre présente un résumé synthétique de cette thèse en français. Dans un premier temps, nous introduisons les problématiques, motivations, objectifs, et contributions de la thèse. Puis, la section 2 présente l'état de l'art sur la gestion de ressources dans un environnement sans fil. Avec les applications multimédias d'aujourd'hui, il est important de prendre en compte la qualité perçue par l'utilisateur. Pour cela, la section 3 présentera différentes façons d'évaluer cette qualité appelée *qualité d'expérience (QdE)* afin de trouver la méthode la plus appropriée. L'accent sera mis sur une technique hybride utilisant l'évaluation pseudo-subjective appelée *PSQA (Pseudo-Subjective Quality Assessment)*. Ensuite, la section 4 décrit comment la QdE peut être déployée dans la gestion des ressources et des exemples de mécanismes orientés QdE sont proposés aussi bien pour le côté réseau que pour le côté utilisateur. Enfin, la section 5 approfondi les résultats obtenus et conclu en rappelant les contributions de ce travail avant de décrire les perspectives ouvertes en donnant quelque orientation de recherche.

Problématique de recherche

Au début des réseaux, les connexions entre machines étaient réalisées via des câbles pour établir ce que nous appelons maintenant un *réseau câblé*. Ce type de réseau offre dorénavant une bande passante et une stabilité élevée, ce qui facilite la gestion des ressources du réseau. Grâce aux progrès technologiques et à un besoin sans cesse croissant d'avoir une connexion permanente, les réseaux sans fil et mobiles se sont de plus en plus développés. De nombreux produits et applications ont été développés pour fonctionner sur ce type de réseau. Un ordinateur personnel d'aujourd'hui peut travailler sur les deux environnements: filaires et sans-fil. Plus précisément, les terminaux mobiles peuvent désormais connecter l'utilisateur à Internet via différents réseaux d'accès simultanément.

Pendant ce temps, les utilisateurs ont un intérêt grandissant pour les applications multimédias. Ce type de trafic croît actuellement de manière considérable sur les réseaux, ce qui change de l'époque où les utilisateurs étaient la plupart des spécialistes du domaine utilisant l'Internet dans la sphère professionnelle. Aujourd'hui le plus grand nombre d'utilisateurs des réseaux sont des non spécialistes plus intéressés par la qualité qu'ils perçoivent et non par les paramètres techniques d'évaluation de la qualité. En conséquence, la qualité proposée doit dorénavant être mesurée en termes de qualité telle que perçue par l'utilisateur plutôt que seulement en termes de paramètres réseaux.

Dans ce contexte, les deux principaux acteurs sont *l'opérateur de réseau* et *l'utilisateur*. Le rôle de l'opérateur est de fournir des services aux utilisateurs via différents réseaux d'accès : l'utilisateur du réseau est alors client des services fournis. On peut remarquer dans ce modèle que les utilisateurs jouent un rôle important puisque leur satisfaction est fondamentale pour l'opérateur. La ressource essentielle qui doit être gérée est *la bande passante*, qui est limitée et variable à cause de sa nature sans fil. Du point de vue de l'opérateur, la bande passante doit être répartie de manière efficace afin de maximiser le nombre d'utilisateurs et par conséquent maximiser les revenus. Les utilisateurs quant à eux veulent choisir le meilleur réseau, c'est-à-dire celui qui fournira la meilleure qualité avec les coûts réduits. Dans une telle situation la gestion des ressources est cruciale, car seuls des mécanismes efficaces peuvent satisfaire les deux parties. Différents facteurs qui compliquent la gestion sont listés ci-dessous.

- Tout d'abord, *la nature du réseau sans-fil* rend la gestion plus difficile. Avec son environnement ouvert, un réseau sans-fil est sensible à de nombreuses perturbations. En conséquence, l'état du réseau varie souvent et la garantie de qualité de service peut devenir une question complexe.

- Un deuxième facteur est *l'augmentation du trafic* en raison du nombre croissant des utilisateurs dans l'Internet. Beaucoup de progrès ont été réalisés et les terminaux sont maintenant abordables pour presque tout le monde. Les réseaux d'accès sont divers et accessibles presque partout, ce qui rend beaucoup plus facile l'obtention d'une connexion. Ce phénomène augmente la difficulté de gestion des ressources puisque l'augmentation du trafic entraîne une augmentation de la congestion et des interférences.

- Un autre facteur important est *le développement croissant des applications multimédias* dans les réseaux sans-fil. Avec ce type d'application, la façon de gérer les ressources est actuellement déterminée par des paramètres techniques. Ce n'est pas optimal puisque de nombreuses applications multimédias génèrent du trafic avec un débit variable. La gestion de qualité en utilisant, par exemple, le paramètre bande passante n'est pas suffisante, surtout dans un environnement sans-fil où les ressources sont rares et l'état du médium radio extrêmement variable.

- Résultant de *la croissance des applications multimédias*, la Qualité de Service (QdS) devient moins importante et la notion de Qualité d'Expérience (QdE), ou parfois appelée expérience utilisateur, est de plus en plus significative pour l'utilisateur du réseau. La QdE révèle la qualité d'un service tel qu'il est perçu par l'utilisateur. Comme l'objectif final de tous les services est la satisfaction des clients, la qualité d'expérience devient donc la préoccupation principale. En effet, les opérateurs de réseaux veulent maximiser leur profit en optimisant l'utilisation des ressources, mais en même temps s'assurer de la fidélité de leurs utilisateurs ce qui résulte directement de leur satisfaction.

- Une grande *variété d'applications* dans les réseaux sans-fil d'aujourd'hui rend la gestion des ressources très difficile à traiter. Différents types d'application (VoIP, streaming vidéo, jeux interactifs, emails, FTP, etc.) ont des exigences différentes en termes de bande passante, délai, gigue, etc. Par conséquence, un traitement approprié est nécessaire pour chaque type d'applications si nous voulons satisfaire les attentes des utilisateurs.

- Enfin, l'*hétérogénéité des réseaux d'accès* ou ce que l'on pourrait appeler "le réseau hétérogène" devient une réalité. Les appareils actuels sont en effet équipés de plusieurs interfaces permettant la connexion à différentes technologies de réseau (Ethernet, Wi-Fi, cellulaire, satellite, etc.), même de manière simultanée. Ces diverses technologies ont des caractéristiques différentes et peuvent être combinées afin de fournir un système hétérogène très puissant permettant à toutes les classes d'applications de trouver le réseau d'accès adéquat. L'arrivée de ce type d'environnement nécessite ainsi un traitement spécial et augmente la complexité du problème de gestion.

Pour résumer, cette thèse se concentre sur les problèmes de gestion des ressources dans les réseaux sans-fil. Les thèmes qui sont traités sont l'allocation de bande passante et la gestion des connexions. Les aspects commerciaux tels que la tarification et le SLA (Service Level Agreement) sont en dehors des champs d'investigation de cette thèse. Dans le cas d'un environnement hétérogène, on suppose que l'opérateur de réseau possède les différentes technologies de réseau nécessaires aux tests. Par ailleurs, les aspects traitant de la sécurité comme l'authentification et l'autorisation ne font pas partie des objectifs de cette thèse. Par conséquent, un serveur de type AAA (authentification, autorisation et comptabilité), est supposé être présent dans le réseau pour gérer tous ces aspects.

Motivations et objectifs

L'objectif principal de cette thèse est de trouver de nouvelles solutions pour la gestion des ressources en tenant compte de tous les aspects décrits ci-dessus. L'environnement auquel l'on s'intéresse concerne les réseaux sans-fil multimédias. Le travail se concentrera sur deux points de vue : celui de l'opérateur de réseau et celui de l'utilisateur du terminal client. Plusieurs solutions seront proposées.

- Pour l'approche *orientée opérateur de réseau*, la maximisation des revenus est sa préoccupation principale. Des informations utilisateurs sont souvent recueillies pour la prise de décision. Des mécanismes sont, par exemple, le contrôle d'admission (*admission control*) qui gère le trafic entrant en choisissant d'accepter ou non chaque nouvelle connexion; l'ajustement du trafic s'effectue à un point d'attachement qui adapte le débit (*rate adaptation*) afin de mieux répondre à l'état actuel du réseau; ou l'ordonnanceur (*packet scheduling*) qui planifie l'allocation de bande passante en fonction de la qualité et/ou de priorité ainsi que de la classe de trafic.

- Pour l'approche *orientée utilisateur*, la décision finale a lieu sur le terminal client. Le profit de l'utilisateur (en termes de prix ou de qualité d'expérience par exemple) est le principal critère de décision, même si certains mécanismes peuvent également prendre en compte les informations provenant de l'environnement réseau. Habituellement, l'approche orientée utilisateur est principalement liée au mécanisme de sélection du réseau d'accès (*network selection*) pour choisir le meilleur réseau lorsqu'il y en a plusieurs d'accessibles. On remarque qu'à part ce mécanisme, le terminal n'a pas beaucoup d'autres contrôles dans le réseau.

Dans cette thèse, on souhaite mettre en avant la notion de qualité d'expérience dans la gestion de ressources. Par définition, la qualité d'expérience est liée à la couche applicative. Cependant, nous pouvons traiter les problèmes liés à QdE également au niveau d'autres couches. Pour la couche applicative, certaines adaptations peuvent être faites sur le terminal utilisateur ou sur le serveur afin d'améliorer la qualité du flux. Cela inclut par exemple une technique comme le changement du taux d'encodage, que le serveur multimédia peut modifier dynamiquement en fonction de l'état du réseau. Au niveau de la couche réseau (IP), la qualité peut être améliorée si nous pouvons contrôler correctement l'état du réseau. Pour la couche liaison (MAC) et la couche physique (PHY), le débit peut être adapté afin de répondre aux conditions physiques du réseau. Dans cette thèse, l'accent sera principalement mis sur les couches IP et MAC, où les contrôles peuvent être exécutés par l'opérateur réseau. En outre, les contrôles possibles au niveau de la couche applicative du côté du terminal utilisateur seront aussi étudiés.

Contributions

Tout d'abord, un état de l'art sur la problématique de la gestion des ressources a été réalisé. Il concerne principalement les applications multimédias avec des exigences élevées auxquelles il est difficile de garantir un haut niveau de qualité de service. Par rapport à cela, des mécanismes de gestion qui visent à satisfaire l'utilisateur et en même temps à optimiser l'utilisation des ressources ont été étudiés. Du côté réseau, des mécanismes orientés QdE comme le contrôle d'admission, l'adaptation de débit, et l'ordonnancement ont été proposés. De même, du côté terminal des mécanismes de sélection de réseau d'accès ont été proposés. La plupart d'entre eux prennent en compte les informations de l'utilisateur et du réseau pour couvrir l'ensemble des critères. Les études ont été menées dans différentes technologies sans-fil (IEEE 802.11 et Cellular Network) à la fois dans des contextes homogènes (i.e. réseau utilisant une seule technologie) et également dans des contextes de réseaux sans-fil hétérogènes (i.e. multi-technologie). Les résultats obtenus démontrent qu'il est possible et utile d'utiliser la qualité d'expérience en tant que métrique pour améliorer la gestion des ressources dans les réseaux mobiles.

La liste des publications concernant ces travaux est présentée ci-dessous:

- [1] K. Piamrat, A. Ksentini, J.-M. Bonnin, and C. Viho. "Radio resource management in emerging heterogeneous wireless networks". Computer Communications, In Press, Corrected Proof, Available Online, Feb. 2010.

- [2] K. Piamrat, K. Singh A. Ksentini, C. Viho, and J.-M. Bonnin. "QoE-aware scheduling for video-streaming in High Speed Downlink Packet Access", IEEE Wireless Communications & Networking Conference (WCNC 2010), 18-21 Apr. 2010.

- [3] K. Piamrat, A. Ksentini, J.-M. Bonnin, and C. Viho. "Q-DRAM: QoE-based dynamic rate adaptation mechanism for multicast in wireless networks". In IEEE Global Telecommunications Conference (GLOBECOM 2009), pages 1-6, 30 Nov. - 4 Dec. 2009.

- [4] K. Piamrat, A. Ksentini, J.-M. Bonnin, and C. Viho. "Rate Adaptation mechanism for Multimedia Multicasting in Wireless Networks", Sixth International Conference on Broadband Communications, Networks, and Systems (Broadnets 09), pages 1-7, Sep. 2009.

- [5] K. Piamrat, C. Viho., J.-M. Bonnin, and A. Ksentini. "Quality of Experience Measurements for Video Streaming over Wireless Networks", In Sixth International Conference on Information Technology: New Generations (ITNG 09), pages 1184 -1189, April 2009.

- [6] K. Piamrat, A. Ksentini, C. Viho, and J.-M. Bonnin. "QoE-Aware Admission Control for Multimedia Applications in IEEE 802.11 Wireless Networks". In IEEE 68th Vehicular Technology Conference (VTC 2008-Fall), pages 1-5, Sep. 2008.

- [7] K. Piamrat, A. Ksentini, C. Viho, and J.-M. Bonnin. "QoE-based network selection for multimedia users in IEEE 802.11 wireless networks". In 33rd IEEE Conference on Local Computer Networks (LCN 2008), pages 388-394, Oct. 2008.

- [8] K. Piamrat, C. Viho, A. Ksentini, and J.-M. Bonnin. QoE-aware Network Selection in Wireless Heterogeneous Networks. In Research Report RR-7282, INRIA, 2010.

2. Etat de l'art

Avec l'évolution des réseaux de nos jours, la qualité devient un facteur critique car ce paramètre fait fonctionner l'économie de plusieurs façons, par exemple via des accords de SLA (Service Level Agreement) ou encore au niveau de la fidélisation des clients. Pour un utilisateur, la qualité finale d'un service est une combinaison de disponibilité, qualité, prix, et rentabilité. En conséquence, la gestion des ressources doit s'effectuer en temps réel et doit prendre en compte la perception de l'utilisateur également appelée la qualité d'expérience.

Selon l'ITU [9], la Qualité d'Expérience (QdE) est l'acceptabilité globale d'une application ou un service, tel qu'il est perçu subjectivement par l'utilisateur final. Elle diffère de la fameuse qualité de service (QdS) à bien des égards. Tout d'abord, la QdE est subjective et se rapporte à la qualité d'un service perçue par l'utilisateur final, alors que la QdS est objective et se rapporte aux états courants du réseau ou du flux de trafic. En d'autres termes, la QdE mesure comment les entités du réseau satisfont les besoins et les attentes de l'utilisateur.

Avant l'avènement des communications multimédias, les paramètres de la QdS étaient suffisants pour évaluer la qualité des services fournis. Toutefois, les applications multimédias se multipliant de plus en plus, et les utilisateurs devenant également de plus en plus expérimentés, la notion de qualité s'est déplacée du niveau réseau au niveau utilisateur. Les mesures existantes ne suffisent plus dorénavant à refléter la perception d'un service que pourrait avoir un utilisateur. Par exemple le taux de perte, un indicateur largement utilisé dans le domaine de la qualité, n'est pas toujours fiable lorsqu'il s'agit de qualité d'expérience. En fait, une perte élevée ne signifie pas automatiquement une mauvaise perception. Si l'expéditeur utilise une technique de prévention comme la FEC (Forward Error Correction), la QdE peut être maintenue à un niveau acceptable malgré des pertes élevées.

Concernant les technologies réseaux, le sans-fil se propage progressivement et a donné naissance aux réseaux multimédia sans-fil ou encore appelés WMN (Wireless Multimedia Networking). Avec ce type de réseau, la distribution de charge doit donc être soigneusement contrôlée afin que la qualité reste acceptable tout en veillant à ce que les opérateurs de réseau ne soient pas non plus pénalisés par une sous-utilisation. En général, pour garantir une bonne perception auprès des usagers, un opérateur IP triple-play à large bande doit toujours veiller à ce que chaque lien principal transporte moins de 50% de sa capacité, cela pour éviter la congestion en cas de défaillance d'un lien redondant.

Dans cette thèse, nous essayons d'éviter une telle approche conservatrice en étudiant les possibilités et les performances de l'utilisation de QdE comme métrique pour la gestion des ressources. Ce nouveau paradigme permettra une meilleure flexibilité, tout en maximisant l'utilisation des débits et en maintenant une perception satisfaisant aux utilisateurs. Nous nous concentrons tout d'abord sur un environnement homogène. Enfin, nous étudierons un environnement hétérogène.

Gestion de ressources

La gestion de ressources est illustrée dans la figure 1. Trois étapes majeures peuvent se distinguer : la surveillance des ressources, la prise de décision et la mise en œuvre de la décision.

- **Surveillance des ressources** - C'est la phase durant laquelle l'information est recueillie. Ces données proviennent des utilisateurs et/ou des réseaux. La collecte d'informations peut varier d'un mécanisme à l'autre et l'information recueillie sera considérée comme une entrée pour la prise de décision. Nous pouvons voir sur la figure 1 que le contrôle des ressources se situe à deux moments différents: avant de se connecter au réseau et après l'établissement de la connexion.

  Le premier contrôle vise à surveiller le réseau et à recueillir des informations pour la toute première décision (*sélection du réseau d'accès* ou *allocation de bande passante*). Puis s'il n'y a pas de solution (c'est-à-dire que les réseaux existants ne correspondent pas aux exigences), l'utilisateur peut avoir à modifier ses exigences (*application adaptation*, cf. la figure 1) afin de trouver le réseau approprié. Si l'adaptation n'est pas possible, l'utilisateur retournera surveiller les ressources et attendra un meilleur état du réseau.

  Le second type de contrôle vise à observer l'état de la connexion en cours pour déclencher une adaptation du réseau lorsqu'un événement se produit, par exemple lorsque l'utilisateur se déplace hors de la cellule courante (mobilité) ou en cas de congestion du réseau. Dans ces situations, une nouvelle décision d'adaptation doit être prise en tenant compte de la nouvelle situation.

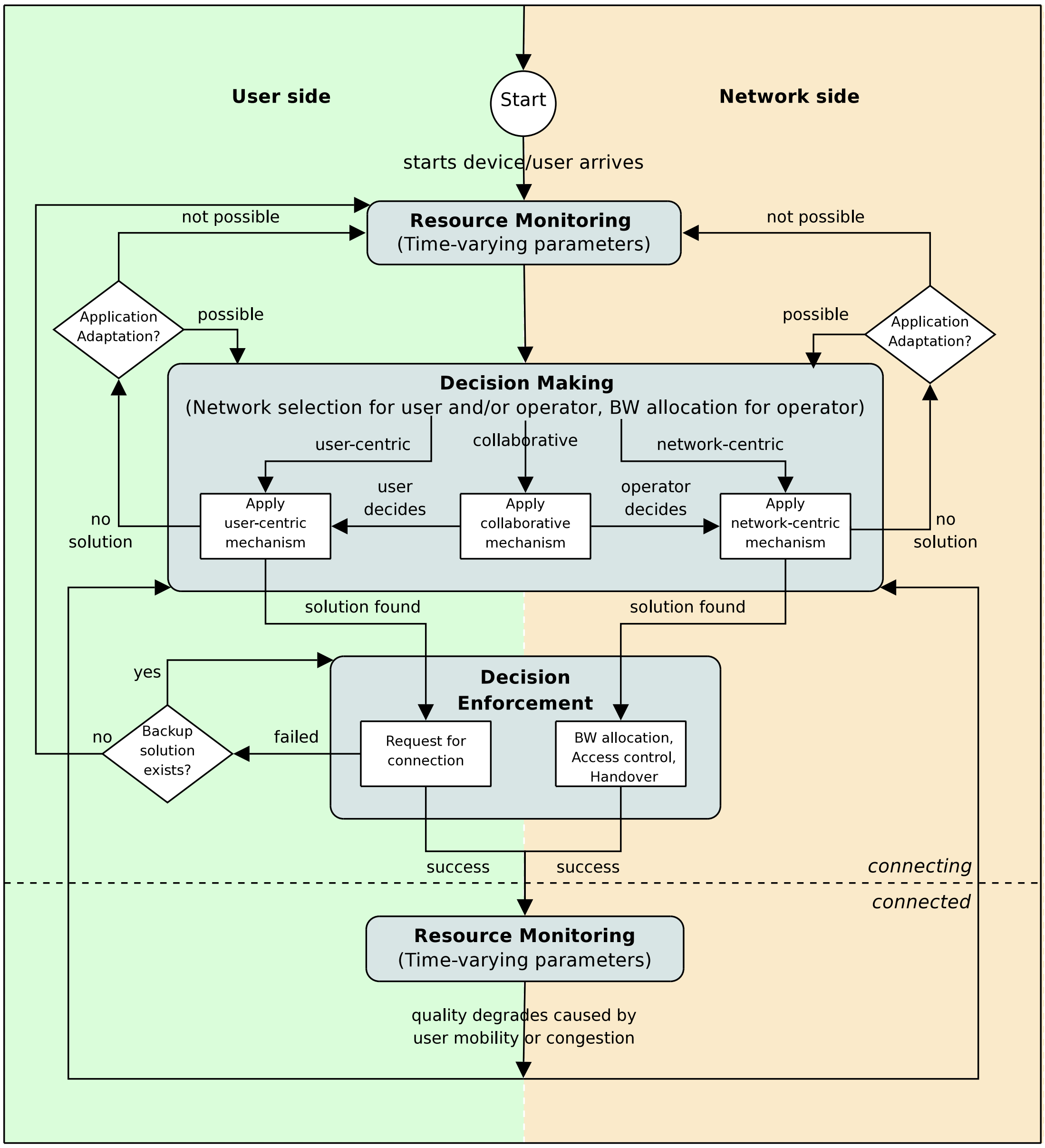


Figure 1: Vision globale de la gestion de ressources dans un réseau hétérogène.

- **Prise de décision -** La plupart du temps, ces décisions sont prises par l'opérateur du réseau (approche centrée réseau) mais elles peuvent également s'effectuer au niveau des terminaux utilisateurs (approche centrée utilisateur), ou encore ces décisions peuvent être le fait d'une collaboration (approche collaborative). Les deux principales décisions à prendre sont le choix du réseau (comment sélectionner le meilleur réseau disponible pour une connexion) et l'allocation de bande passante (la façon de répartir la bande passante des différents réseaux aux utilisateurs). Dans un réseau hétérogène, la bande passante allouée à une connexion donnée peut impliquer plusieurs réseaux d'accès. Dans ce cas la ressource est appelée *ressource commune*. On peut noter que, dans l'approche collaborative, la décision finale sera prise par un seul des deux acteurs (soit l'utilisateur ou l'opérateur), puis les étapes suivantes correspondent à deux approches centrées sur l'utilisateur ou du côté de l'opérateur réseau. La prise de décision représente le cœur du problème; par conséquent, elle sera abordée plus en détails. La classification des mécanismes se fera en fonction de décideur final.

- **Mise en œuvre de décision -** C'est la phase dans laquelle les décisions sont exécutées. Dans l'approche centrée utilisateur, une demande de connexion au réseau sélectionné est effectuée, si cette demande échoue on cherchera alors à contacter la deuxième meilleure possibilité et ainsi de suite. Si aucune des décisions ne peut être mise en œuvre, une nouvelle phase de surveillance du réseau sera effectuée. Cette situation peut se produire par exemple lorsque le réseau refuse une demande entrante afin de protéger la performance globale des utilisateurs en cours. Dans les approches centrées réseau, la sélection du réseau est appliquée en utilisant des mécanismes de contrôle d'admission pour filtrer l'accès aux réseaux en fonction de la décision rendue à l'étape précédente. En outre, la décision de déplacer les utilisateurs vers un autre réseau d'une technologie identique ou différente est exécutée par des mécanismes comme le "handover vertical et horizontal". Pour l'allocation de bande passante, l'opérateur distribue la bande passante en fonction de la décision prise.

Mécanismes de décision

Un tableau récapitulatif des approches récemment proposées dans la littérature est présenté ci-dessous. Pour plus d'information, voir le chapitre 2.

Techniques	**Parameters**	**Procedure**	**Output**	**Approach**	**Joint allocation**
SLP	Allocation, demand, underutilization, and rejection	1-association of predetermined probability to demands 2-variable formulation 3-SLP statement	Allocation in each network	Network-centric	Yes
Game Theory	Available bandwidths in each network	1-determine offered bandwidths 2-compute Shapley value 3-verify core	Bandwidth allocation	Network-centric	Yes
Degradation Utility	Released bandwidth and lost reward point	1-compute ratio of released - bandwidth & loss reward point for each connection 2-find maximum	Connection that gives maximum utility	Network-centric	No
AHP & GRA	User's requirements and network conditions	1-AHP of user's requirements 2-GRA of network conditions 3-compute GRC	Network rank by GRC	User-centric	No
Consumer Surplus	Utility and cost	1-compute the difference between utility and cost for each network 2-find maximum	Network that gives maximum benefit	User-centric	No
Profit function	Bandwidth gain and handoff cost	1-compute the difference between gain and cost for each network 2-find maximum	Most appropriate network for handoff	User-centric	No
FLC	Network data rate, SNR, application - required data rate	1-fuzzification 2-fuzzy inference 3-defuzzification	Fitness rank of each network	Collaborative	No
Objective function	Quality and policy indicators	1-compute sum of (inputs$\times$weights) for each network 2-find maximum	Allocation of services to APs and terminals	Collaborative	Yes
TOPSIS	QoS parameters and traffic class	1-format data into normalized matrix 2-compute data$\times$their weights 3-compute ideal points (+/-) and distances from ideal points 4-select the best solution	Best path for flow distribution	Collaborative	No

Table 1: Résumé des approches investiguées.

3. Mesurer la qualité d'expérience

Avant d'être capable de déployer la QdE dans la gestion de réseau, un outil de mesure approprié est nécessaire. Pour une meilleure compréhension de la QdE, cette section donne un aperçu des différentes approches utilisées pour mesurer la QdE, allant de l'approche subjective traditionnelle aux approches objectives et hybrides respectivement. A la fin de la section, leurs performances sont comparées afin de choisir la méthode la plus appropriée pour l'étude.

Approche subjective

Il est évident que la façon la plus précise pour mesurer la qualité perçue est l'évaluation subjective des utilisateurs car il n'y a pas d'autre indicateur de qualité perçue meilleure que celui évalué par l'homme. Cela consiste en la construction d'un panel d'observateurs humains qui vont donc évaluer des séquences de médias sur ce critère. Le résultat de cette évaluation est donné en termes de note moyenne d'opinion (MOS - Mean Opinion Score), sur une échelle à cinq niveaux (présentée dans le tableau 2).

Table 2: Définition du MOS et conversion possible de PSNR.

MOS	**Qualité**	**Altération**	PSNR
5	Excellent	Imperceptible	> 37
4	Bien	Perceptible mais non gênant	31-37
3	Acceptable	Un peu gênant	25-31
2	Mauvais	gênant	20-25
1	Très mauvais	Très gênant	< 25

Des méthodes normalisées pour la campagne des évaluations subjectives de la qualité vidéo sont données dans ITU-R BT.50011 [10] avec plusieurs variantes. Pour une campagne d'évaluations subjectives appropriée, il est nécessaire de choisir parmi les différentes options disponibles celles qui conviennent le mieux aux objectifs et aux contextes des problèmes. Dans ce chapitre, nous nous intéressons à la méthode *Stimulus Simple (SS)*, dans laquelle les séquences vidéos sont présentées une par une et où l'évaluateur fournit un score pour chacune d'entre elles (comme le montre la figure 2). Le score final de chaque séquence vidéo est la moyenne des notes de tous les observateurs, à l'exclusion des notes extrêmes (filtrées par un filtre statistique). D'autres variantes de tests subjectifs sont à peu près similaires, les changements pouvant alors concerner l'échelle d'évaluation, la vidéo de référence, la longueur de la séquence vidéo, le nombre de la vidéo, ou encore le nombre d'observateurs.

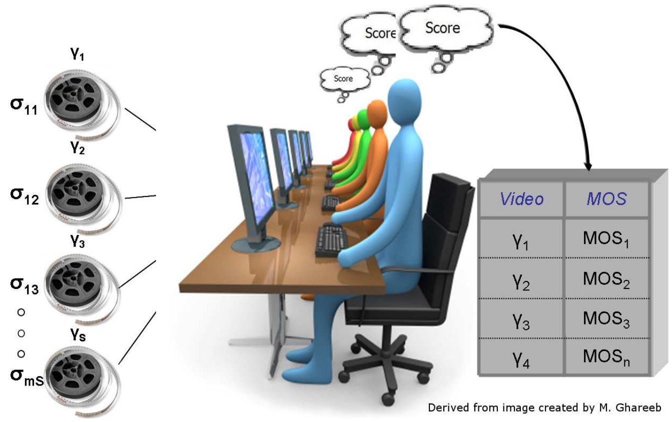


Figure 2: Campagne de mesures subjectives.

Bien que cette approche subjective soit la plus précise, sa mise en place est très coûteuse en termes de temps et de main-d'œuvre (à cause de la campagne subjective). La procédure d'évaluation est très complexe et a des exigences strictes. Par conséquent, on ne peut guère l'utiliser dans une mesure automatique ou pour des outils de suivi en temps réel.

Approche objective

Puisque l'approche subjective est difficile à mettre en œuvre, une approche objective a donc été proposée. Celle-ci utilise des algorithmes, des formules et des mesures de QdS d'un flux donné via des paramètres techniques qui peuvent facilement être collectés dans le réseau. Parmi les méthodes objectives, nous nous sommes intéressés au *Peak Signal to Noise Ratio (PSNR)*, qui est une méthode simple et couramment utilisée pour évaluer la qualité des vidéos. En effet, PSNR [11] est le rapport entre la puissance maximale possible d'un signal et la puissance du bruit qui affecte la fidélité de sa représentation. Il est défini par l'erreur quadratique moyenne (Mean Squared Error-MSE) entre une trame originale o et une trame déformée d comme suit:

$$MSE = \frac{1}{mn} \sum_{m=1}^{m} \sum_{n=1}^{n} |o(m,n) - d(m,n)|^2. \tag{1}$$

Chaque trame a *MxN* pixels, *o(m,n)* et *d(m,n)* sont les pixels de luminance de la position *(m,n)* dans la trame. Le PSNR représente le rapport logarithmique entre la valeur maximale d'un signal et le bruit de fond (MSE). Si la valeur de luminance maximale dans la trame est L (lorsque les pixels sont représentés à l'aide de 8 bits par échantillon, L = 255), on a alors :

$$PSNR = 10 \log \frac{255^2}{MSE}. \tag{2}$$

Il peut être remarqué que le PSNR ne peut être calculé que lorsque l'image est reconstruite au niveau du récepteur. Par conséquent, il est impossible de l'utiliser dans les mécanismes en temps réel. En outre, la conversion de PSNR en MOS est toujours discutable. Si le PSNR est utile pour mesurer la proximité de l'image compressée par rapport à l'original au niveau du signal, il ne prend pas en compte la qualité visuelle de la reconstruction et ne peut être considéré comme une mesure objective fiable de la qualité visuelle d'une image. Cependant, d'après Gross et al. [12], les mappages heuristiques possibles de PSNR vers MOS existent; ils sont présentés dans le tableau 2.

Approche hybride

En dehors des deux approches précédentes, une approche hybride propose un compromis. De nombreuses méthodes ont été proposées pour mesurer la QdE dans des applications de voix sur IP ou VoIP (e.g. E-Model ITU G.107 [13], PSQM et MNB ITU P.861 [14], ou PESQ ITU P. 862 [15]) mais très peu existent pour l'application streaming de vidéo. Dans ce document, nous nous intéressons à *PSQA (Pseudo-Subjective Quality Assessment)* [16], une évaluation pseudo-subjective de la qualité, qui fournit l'évaluation de la QdE en temps réel avec des résultats similaires à la perception humaine. PSQA est basé sur l'apprentissage statistique à l'aide de réseaux de neurones aléatoires (Random Neural Network - RNN) [17]. Il est hybride dans le sens où il y a tout de même une évaluation subjective dans la méthodologie. Toutefois, cette évaluation subjective peut n'être réalisée qu'une seule fois et utilisée plusieurs fois. Avant d'être en mesure d'utiliser PSQA dans les évaluations en temps réel, trois étapes doivent être faites. Les détails de chaque étape peuvent varier selon les contextes. La méthodologie pour l'application streaming de vidéo est expliquée ici.

1-Facteurs de qualité et construction de la base de données des vidéos déformées

Dans une première étape, nous sélectionnons un ensemble de facteurs de qualité qui ont un impact significatif sur la qualité, tels que le codec, la bande passante, la perte, le délai, ou la gigue ainsi que leurs intervalles de valeurs. Un ensemble de paramètres avec des valeurs données est appelé une *configuration*. Une base de données des vidéos déformées est générée en faisant varier des configurations représentatives. La mise en œuvre de cette étape pourrait être faite par des expériences sur une plateforme réelle, un émulateur de réseau ou un simulateur de réseau.

2 - Evaluation de la qualité subjective

Dans la deuxième étape, les configurations choisies précédemment sont utilisées dans une campagne d'évaluation subjective. La méthode *Stimulus Simple (SS)* est utilisée et un groupe d'observateurs humains évalue les vidéos déformées comme illustré sur la figure 3. Puis le MOS est calculé de la même manière que dans l'approche subjective. Les mappages de configurations et du MOS correspondants sont stockés dans deux bases de données séparées, l'un pour l'entraînement et l'autre pour la validation.

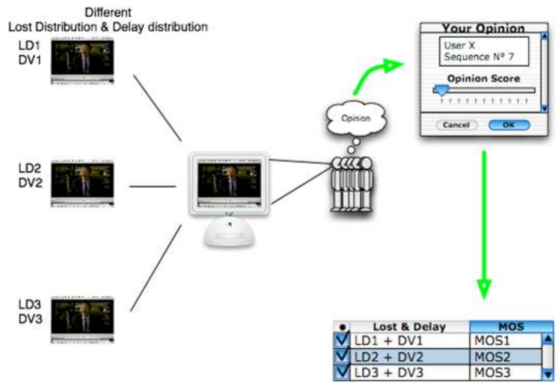
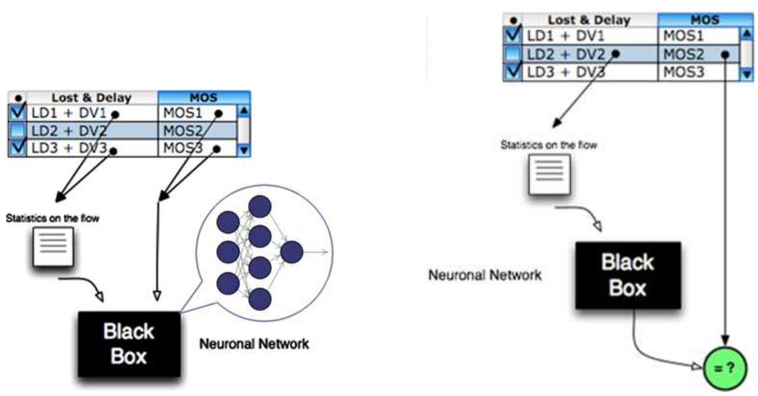


Figure 3: Campagne d'évaluation subjective.

3 - Apprentissage du comportement de qualité avec RNN

Dans cette étape, le RNN apprend les mappages des configurations et les scores tels que défini dans la base de données d'apprentissage. Une fois qu'il a été entraîné, nous obtenons une fonction $f()$ qui permet de mapper toutes les valeurs possibles des paramètres en MOS. Le RNN est validé en comparant la valeur donnée par cette fonction $f()$ au point correspondant à chaque configuration dans la base de données de validation (que le RNN n'a pas vu avant). Si les valeurs sont assez proches, le RNN est validé. Sinon, les configurations choisies doivent être réexaminées et les étapes 1 à 3 doivent être répétées jusqu'à ce que le RNN soit validé.



(a) Apprentissage (b) Validation

Figure 4: Apprentissage du comportement de qualité avec RNN.

Une fois que le RNN a été validé, PSQA est facile à utiliser. Pour obtenir un score instantané à l'instant t, il suffit de mesurer la valeur des paramètres affectant la qualité à l'instant t et de les donner au RNN, qui renvoie instantanément la valeur du MOS.

Comparaison de performance

Une expérience, en utilisant l'application streaming de vidéo sur WLAN (Wireless Local Area Network), a été menée sur l'évaluation de la QdE avec les trois méthodes décrites précédemment. La figure 5 illustre la comparaison entre PSQA (hybride) et PSNR (objectif) en référence à une méthode "Single Stimulus" (subjective) dans des conditions variables du réseau. Seul le taux de perte est utilisé dans nos tests car il est le facteur le plus important de la qualité, les autres paramètres réseaux comme le retard, la gigue, le débit sont affectés en fonction des pertes dans le réseau. Nous ne testons pas au delà d'un taux de perte de 10% parce que des pertes plus élevées donneront toutes des résultats de qualité "Mauvaise". Pour générer des pertes réalistes (avec rafale), un modèle Gilbert simplifié [18] est utilisé.

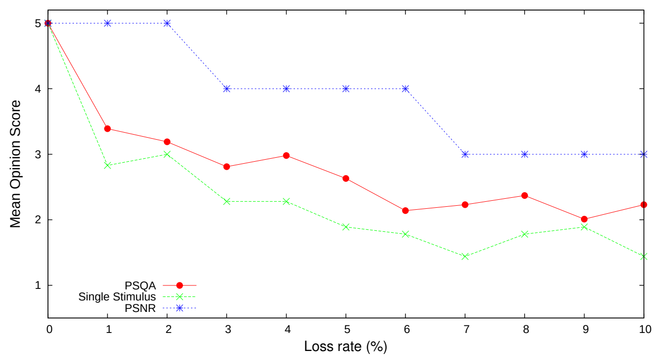


Figure 5: Comparaison entre *PSQA* et *PSNR* avec pour référence *Single Stimulus*.

On peut voir dans la figure 5 que PSQA surpasse PSNR en donnant des approximations plus proches de la méthode subjective dans presque tous les cas. Comme il est difficile et coûteux d'utiliser la méthode subjective en temps réel, le résultat obtenu montre que PSQA est une solution intéressante. C'est elle qui a donc été choisi comme outil d'évaluation de la QdE dans la suite de nos études.

4. Gestion de ressources orientée QdE

Nous avons vu que les paramètres QdS existants ne suffisent pas pour gérer les réseaux multimédia sans-fil d'aujourd'hui. En effet, les paramètres de la QdS sont moins significatifs pour les clients qui ne sont pas vraiment préoccupés par le taux de perte, le retard ou la gigue, mais beaucoup plus par la perception finale de l'utilisateur. Par conséquent, cette section décrit comment la QdE peut être déployée dans les mécanismes de gestion des ressources en donnant quelques exemples. Tout d'abord, une utilisation de PSQA pour la mesure de la QdE est présentée.

4.1 Utilisation de PSQA dans la mesure de QdE

La mesure de la QdE peut s'effectuer à différents endroits : au niveau du terminal utilisateur ou au niveau des équipements réseaux. L'avantage d'une mesure sur le terminal est sa précision, car la mesure est située au niveau du terminal lui-même et les informations peuvent être collectées facilement. D'autre part, PSQA peut être exécuté au niveau d'un routeur ou d'un point d'attachement pour être en mesure de réagir directement à la situation actuelle.

Pour tous les cas d'utilisation présentés dans ce document, deux versions de PSQA ont été entrainées et validées pour une application streaming de vidéo dans un environnement sans-fil. La version 1 concerne des facteurs de qualité au niveau IP. Une configuration se compose de taux de perte et de la taille moyenne des rafales de pertes. Ce dernier paramètre est essentiel parce qu'il est démontré par de nombreuses personnes que l'homme préfère généralement la perte en rafale à la perte isolée. Cela s'explique par le fait que les pertes de paquets en rafale conduisent à un taux de perte de trames applicatives inférieures à celui causé par des pertes de paquets isolés [19]. En outre, plus la longueur de rafale est grande, plus réduite est la durée de la vidéo déformée [20]. La version 2 de PSQA concerne quant à elle des facteurs au niveau applicatif[1] : une configuration se compose de taux de perte de trame I/P/B et de la taille moyenne des rafales de pertes de trame I.

Le simulateur de réseau NS-2 est utilisé pour simuler tous les cas d'utilisation, (NS-2.28 et 2.29 [21] pour WLAN, EURANE [22] pour UMTS, et NIST [23] pour HWN). Les versions de NS-2 ont été modifiées afin d'être capables de transmettre des séquences réelles de vidéo. Le module PSQA a été également intégré dans NS-2.

[1]Au niveau applicatif, la vidéo est composée de trois types de trames (I,P,B). Les trames I étant les trames de références et donc les plus importantes pour reconstruire la vidéo.

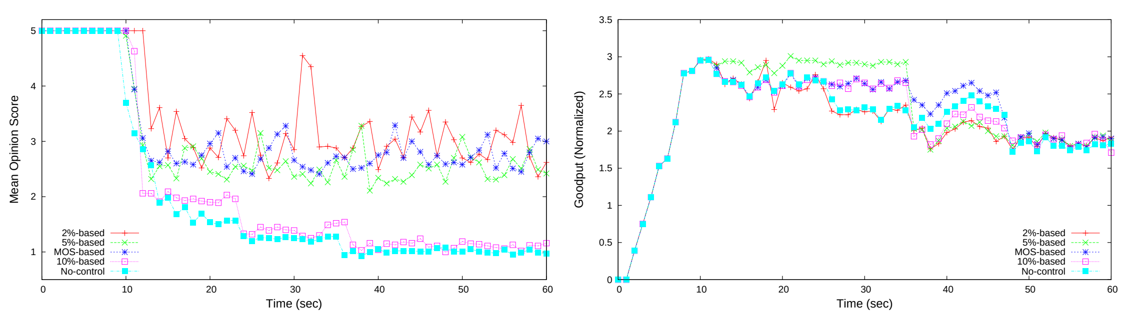
4.2 QdE pour la gestion du côté réseaux

Dans cette section, des exemples de mécanismes de contrôle centré réseaux sont présentés. Ils concernent le contrôle d'admission et l'adaptation de débit multicast dans le réseau Wi-Fi et l'ordonnancement dans le réseau UMTS.

- **Contrôle d'admission**

Dans les réseaux Wi-Fi, l'utilisateur se connecte à Internet via un point d'accès. Comme cette technologie se répand de plus en plus, le nombre d'utilisateurs augmente radicalement et la densité du trafic de chaque zone de couverture s'accroît. L'utilisation généralisée des réseaux sans-fil a mis en évidence un problème de congestion. En outre, l'émergence des applications multimédias accentue encore plus ce problème.

Ainsi, ce premier exemple présente un mécanisme de contrôle d'admission basé sur la QdE perçue par les utilisateurs, appelé "MOS-based" dans la figure 6. Il peut être réalisé en refusant toute nouvelle connexion tant que le MOS des connexions en cours se situe en dessous d'un certain niveau [6]. Dans cette stratégie, le point d'accès surveille le niveau de MOS des utilisateurs courants. Le MOS global du réseau est calculé en prenant la moyenne des notes de toutes les connexions actives. Si cette valeur est supérieure à un seuil, qui est égale à la limite (MOS souhaité) plus la marge de dégradation, une nouvelle connexion peut être acceptée, sinon la nouvelle connexion est rejetée. Cette stratégie est comparée, en termes de satisfaction des utilisateurs (qualité d'expérience) et d'optimisation du réseau (débit utile[2]), avec l'approche basée sur le taux de perte. Avec une telle approche, le point d'accès cesse l'admission d'un nouveau flux lorsque le taux de perte des connexions en cours atteint un certain pourcentage (2%, 5% et 10% dans cet exemple).



(a) Qualité d'expérience (b) Débit utile global

Figure 6: Comparaison des performances entre les différentes approches.

Dans le scénario, à chaque seconde un nouvel utilisateur arrive et le point d'accès prend une décision conformément à la stratégie basée sur l'expérience utilisateur. Le

[2]Le débit utile (ou Goodput en anglais) est, en fait, le débit au niveau applicatif. Il représente le nombre de bits utiles par unité de temps transmis par le réseau à partir d'une source vers une destination. Pour mesurer le débit utile dans NS-2, le nombre de bits reçus avec succès à chaque station est calculé.

Table 3: Résumé des performances.

Mécanismes	Max. débit utilisé	Connexions admis	MOS moyenne
2% based	3.6 Mbps	10 flux	3.62
MOS based	**4.32 Mbps**	**12 flux**	**3.35**
5% based	3.96 Mbps	11 flux	3.19
10% based	4.68 Mbps	13 flux	2.17
Non-control	7.2 Mbps	20 flux	2.06

tableau 3 compare les performances de toutes les approches. On peut remarquer que la stratégie "MOS-based" permet d'admettre plus de flux, tout en maintenant le MOS à un niveau raisonnable. En outre, on peut remarquer d'après la figure 6(a) que la performance de cette stratégie surpasse une approche sans contrôle ainsi qu'une approche basée sur un taux de perte de 10%. L'explication est la suivante. Dans le cas sans contrôle, le contrôle d'admission n'existe pas et le réseau accepte constamment les nouveaux flux. Ce qui mène à une congestion et donc à une mauvaise qualité. Pour l'approche 10%, on remarque que fixer un taux limite de perte à 10% induit une dégradation inacceptable pour des utilisateurs d'une application multimédia telle que le streaming vidéo. La stratégie "MOS-based" a des performances légèrement meilleures que celle basée sur 5% de perte qui est, en général, un taux de perte limite au delà duquel la qualité n'est plus acceptable. La stratégie "MOS-based" obtient à certains moments des meilleurs résultats par rapport à une approche basée sur 2%, mais globalement elle est perdante. La bonne performance de l'approche 2% a une contrepartie, qui est une sous-utilisation de la bande passante comme on peut le voir dans la figure 6(b) et le tableau 3. En effet, une approche qui utilise 2% de perte comme limite d'admission est trop prudente : en conséquence le débit utile de ce mécanisme est plus bas et le point d'accès en utilisant ce mécanisme admet un moins de flux.

- **Adaptation du débit Multicast**

Pour ce mécanisme, un environnement sans-fil multicast sera étudié. Cet environnent est avantageux pour la consommation de bande passante car un paquet n'est envoyé qu'une fois pour atteindre tous les destinataires (clients dans le groupe multicast). Toutefois, avec du Wi-Fi, les paquets multicast sont envoyés avec le débit (modulation) le plus bas, ce qui se traduit par une baisse de la capacité de transmission en raison de l'occupation plus longue du canal.

Pour résoudre ce problème, plusieurs mécanismes ont été proposés. Ils s'appuient sur la possibilité qu'offre le réseau Wi-Fi de transmettre des donnés à des débits différents. Contrairement à d'autres protocoles existants (RAM [24], ARSM [25], LM-ARF [26]) qui utilisent un seuil statique afin de décider quand il faut changer le débit, Q-DRAM (ou "QoE-based Dynamic Rate adaptation Mechanism") utilise une approche avec seuil dynamique [3]. Selon les informations des clients sur la qualité d'expérience, le point d'accès[3] adapte le débit de multicast de la manière suivante: (i) lorsque l'expérience utilisateur est mauvaise, le point d'accès réduit le débit; (ii) lorsque l'expérience utilisateur est bonne, le point d'accès incrémente le débit en fonction du backoff exponentiel binaire (c'est-à-dire, si l'état du réseau devient mauvais (échecs consécutifs), le point d'accès attend deux fois plus longtemps avant de tenter d'augmenter le débit) comme présenté dans la figure 7.

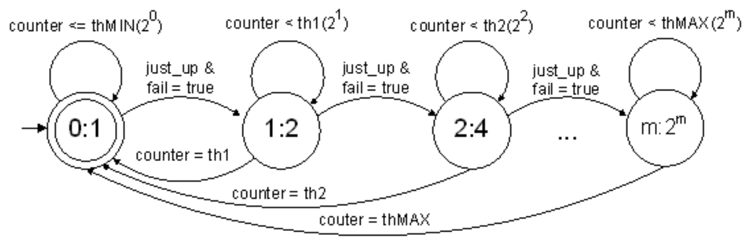


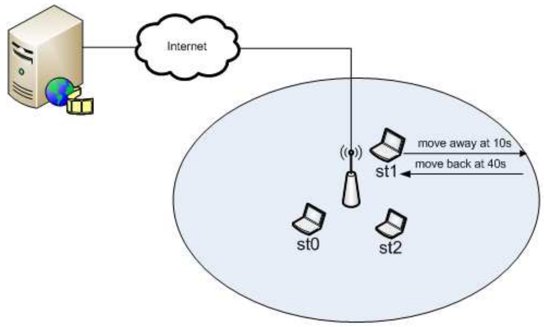
Figure 7: Mécanisme de backoff dans Q-DRAM.

Dans le scénario de la figure 8(a), les clients multicast sont placés autour du point d'accès. Pour générer une erreur de canal (BER-Bit Error Rate), une station est en mouvement durant la période allant de la 15ème à la 45ème seconde. Q-DRAM est comparé avec trois mécanismes: une modification du débit basée sur PSNR (SARM ou "SNR-based Auto Rate for Multicast" [27])), un débit maximal pour avoir la plus grande utilisation de bande passante (11M) et un débit minimal pour avoir le moins de pertes dues au canal (1M).

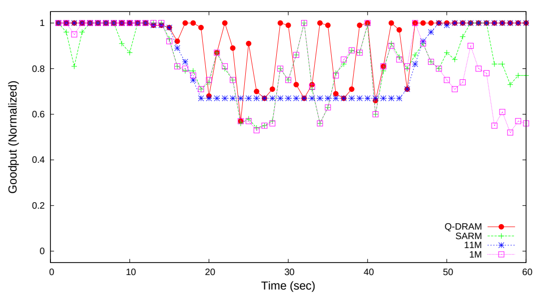
La figure 8(b) illustre la moyenne du débit utile de toutes les stations obtenues pour chaque mécanisme. Il est à noter que le débit utile est normalisé selon le taux d'encodage de la vidéo (le résultat présenté est le débit divisé par le taux d'encodage courant) ainsi les résultats obtenus sont dans l'intervalle [0:1]. La figure 8(b) montre que Q-DRAM fournit un débit utile plus élevé. Plus important encore, son débit utile est significativement plus élevé que tous les autres pendant le mouvement du nœud. En outre, il peut être remarqué que le débit utile est le plus bas lors de la transmission à 1 Mbps, cela est dû à une sous-utilisation de bande passante lors de transmission à faible débit. Le détail de cette anomalie de performance est expliqué dans [28]. En utilisant le taux maximum (11 Mbps) le débit utile est élevé au début et à la fin. Pourtant, lorsque la distance augmente à cause de la mobilité, l'état de canal dégrade (à cause de BER élevé) et cette stratégie a alors une très mauvaise performance. En général, SARM a des résultats légèrement meilleurs que le taux de base (1 Mbps), malgré tout il n'y a pas d'amélioration au cours de la mobilité.

Nous avons observé des fluctuations dans le Q-DRAM au cours de la mobilité car il tente d'augmenter le débit dès qu'il détecte une bonne condition de canal, ceci afin d'obtenir le meilleur débit possible. Malgré ces fluctuations, Q-DRAM surpasse

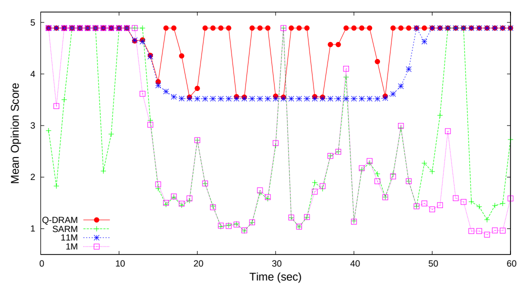
[3]Une abréviation AP (Access Point) est utilisé pour point d'accès.



(a) Scénario et topologie de réseau



(b) Débit utile global



(c) Qualité globale

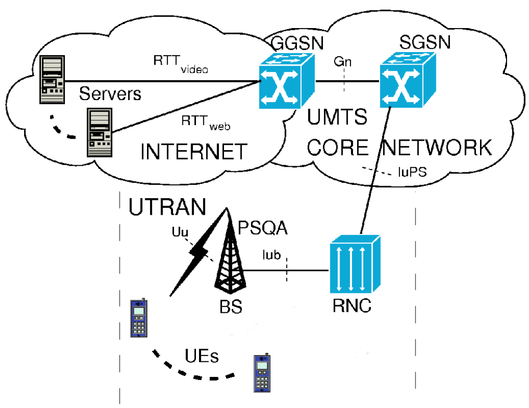
Figure 8: Comparaison des performances de différentes approches.

encore les autres régimes au cours de cette période. La figure 8(c) illustre la satisfaction globale des utilisateurs par le biais du MOS moyen de toutes les stations. Puisque Q-DRAM utilise la qualité d'expérience comme indicateur, il obtient une excellente performance dans la QdE : en effet son MOS moyen est d'au moins 3,5 pendant la session.

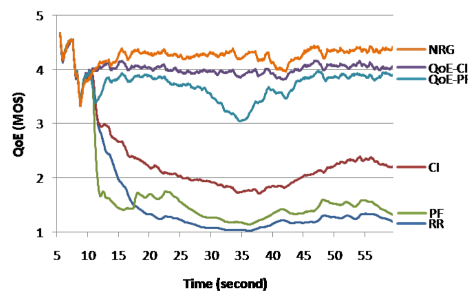
- **Ordonnancement dans l'UMTS**

Regardons cette fois une autre technologie sans-fil, ou ce que l'on peut appeler réseaux cellulaires ou encore réseaux mobiles. Cet exemple concerne UMTS (Universal Mobile Telecommunications System). Avec l'amélioration d'une nouvelle méthode d'accès HSDPA (High Speed Downlink Packet Access), il peut fournir plus de bande passante et assurer une plus large gamme de services y compris les applications multimédia. En UMTS, les différentes catégories de trafic sont précisées ainsi que leurs caractéristiques. Ainsi le trafic "Best effort" a été spécifié avec une basse priorité, car il a moins de contraintes sur la qualité. D'autre part, un trafic multimédia en temps réel comme le streaming de vidéo est plus sensible aux variations de l'état du réseau. Par conséquent, un traitement spécial (par exemple ordonnanceur orienté QdS ou QdE) est nécessaire afin de parvenir à la satisfaction des utilisateurs. D'après la littérature, la plupart des mécanismes d'ordonnancement ne tiennent principalement compte que de la qualité du signal et de l'équité mais ne considèrent pas la perception des utilisateurs.

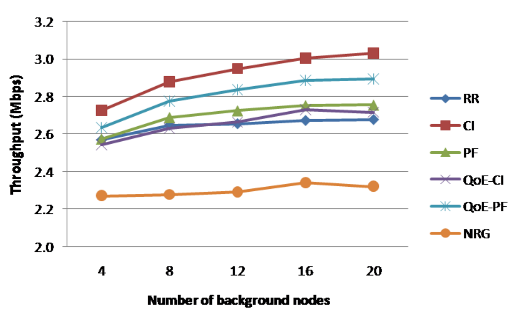
Dans cet exemple, un ordonnanceur orienté QdE [2] sera présenté. Il prend en compte la qualité d'expérience lors des décisions d'ordonnancement. L'idée principale est de donner la priorité aux utilisateurs de streaming vidéo, qui ont plus de contraintes en termes de qualité. Pour cela, un coefficient est attribué à chaque utilisateur. Ce coefficient, d'une façon analogue à la fonction de barrière [29], est alors multiplié avec l'indice de priorité utilisé dans les mécanismes d'ordonnancement traditionnels. L'ordonnanceur orienté QdE différencie le calcul du coefficient des clients vidéo et celui des clients best-effort de la manière suivante : si le MOS des utilisateurs vidéos est inférieur à un seuil spécifique, l'ordonnanceur augmente le coefficient des utilisateurs vidéo et diminue ceux des utilisateurs best-effort. Avec ce procédé les utilisateurs vidéo auront plus de chances d'obtenir une transmission dans le prochain slot.



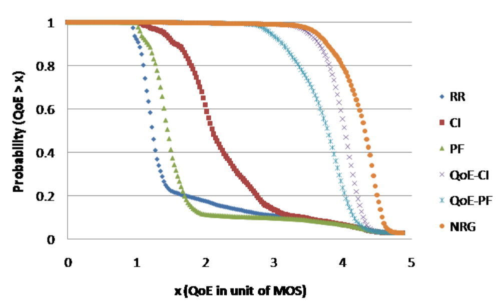
(a) Topologie de réseau



(b) QdE avec différent seuils



(c) Débit avec différent nombres de flux best-effort



(d) Equité

Figure 9: Comparaison des performances de différentes méthodes.

La topologie du réseau est présentée dans la figure 9(a). Dans un scénario de base, il y a 4 nœuds vidéo et 8 nœuds best-effort dans la topologie et leur distance maximale par rapport à la BS est de 300 mètres. L'ordonnanceur orienté QdE est comparé aux approches traditionnelles, à savoir le Round Robin (RR), Maximum Carrier-to-Interference Ratio (CI), Proportional Fair (PF), et l'ordonnanceur orienté

QdS (Normalized Rate Guarantee-NRG [30])). On peut constater dans la figure 9(b) que l'approche orientée QdE a atteint une bonne performance en termes de satisfaction des utilisateurs. Les MOS des utilisateurs de vidéo sont plus élevés que pour les autres approches traditionnelles, mais légèrement inférieurs à celles données par NRG, qui privilégie encore plus les utilisateurs vidéo par rapport aux utilisateurs best-effort. Toutefois, lorsque le nombre de nœuds best-effort augmente dans la figure 9(c), le débit de NRG est très mauvais car il donne trop de créneaux aux utilisateurs de vidéo et pas assez aux utilisateurs best-effort. Finalement, une comparaison d'équité présentée dans la figure 9(d) montre que les ordonnanceurs QdE et NRG sont équitables pour les utilisateurs de vidéo puisqu'ils permettent d'obtenir de bons QdE pour environ 80% des utilisateurs.

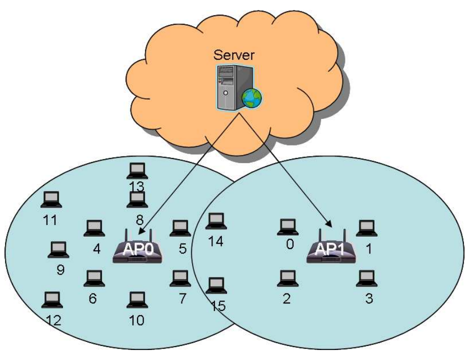
4.3 QdE dans la gestion du côté terminal

Dans cette section, des mécanismes de contrôle du côté terminal utilisateur sont présentés. Ils concernent principalement le mécanisme de sélection de réseau d'accès aussi bien dans l'environnement homogène que dans l'environnement hétérogène.

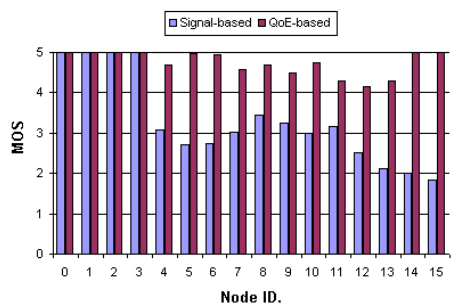
- **Sélection du réseau d'accès - Environnement homogène**

Le réseau Wi-Fi devenant de plus en plus populaire, il y a donc de plus en plus de points d'accès, souvent situés dans une zone géographique très proche que l'on appelle "hotspot". Les utilisateurs doivent pouvoir choisir le réseau qui fournit le meilleur service pour son application. La qualité doit être satisfaite au niveau utilisateur et la performance globale doit être maintenue au niveau du réseau : c'est à dire avoir une répartition équilibrée de la charge entre les points d'accès. Pour cela, un mécanisme de sélection de point d'accès Wi-Fi [7] est proposé dans cet exemple. Il est centré utilisateur et fonctionne avec l'assistance du réseau. En offrant aux utilisateurs des informations pertinentes sur l'état du réseau, ce mécanisme donne un compromis entre la satisfaction des utilisateurs et le rendement de l'opérateur réseau. Pour cela, le point d'accès dans ce système envoie le MOS actuellement perçu par les connexions courantes. Ensuite, les nouveaux clients peuvent décider de se connecter au meilleur réseau disponible. Ceci peut être réalisé en intégrant une note MOS dans les trames "Beacon" et "Probe Response" qui seront envoyées par le point d'accès. Lorsque les utilisateurs passifs reçoivent des "Beacon", ils recevront également le MOS de tous les réseaux disponibles. De même, lorsque l'utilisateur actif envoie une "Probe Request", ils recevront en retour le "Probe Response" avec le MOS. Ce mécanisme basé sur la qualité d'expérience (ou "MOS-based") est comparé avec l'approche par défaut actuellement utilisée dans les terminaux et basée sur des indicateurs de qualité du signal radio. Le scénario est illustré dans la figure 10(a) où les nouveaux utilisateurs ST14 et ST15 décident à quel point d'accès (AP) ils demandent une connexion. Le point d'accès le plus proche de ces deux nœuds est AP0 qui est très chargé. Dans un mécanisme basé sur la qualité du signal, les deux nœuds choisiront AP0 en raison d'un meilleur rapport signal/bruit. Par contre avec le mécanisme "MOS-based", ils trouveront que le MOS d'AP0 est inférieur à celui d'AP1, et ils vont donc préférer se connecter à AP1. La figure 10(b) présente le MOS de chaque nœud, on peut constater que l'approche "MOS-based" surpasse celui basé sur la qualité du signal. La différence de qualité obtenue est de trois niveaux : une amélioration de niveau mauvais à excellent est observée avec ST14 et ST15. Une augmentation importante de la qualité est également illustrée dans toutes les autres stations du réseau d'AP0. En outre, le MOS moyen dans ce réseau présenté dans la figure 10(c) est plus élevé tout le long de la session.

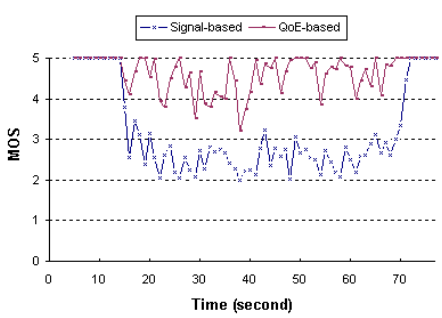
Les charges de chaque réseau d'accès sont illustrées dans la figure 10(d) où on peut remarquer que l'approche orientée QdE donne également de meilleures performances en termes de répartition de charge. La différence entre les charges du réseau représenté par AP0 et AP1 est deux fois plus petite que celle du mécanisme basé sur le niveau du signal reçu. Une bonne performance est automatiquement obtenue avec la sélection du réseau basée sur la QdE puisque les utilisateurs préfèrent le réseau avec une bonne note de MOS, généralement à faible charge.



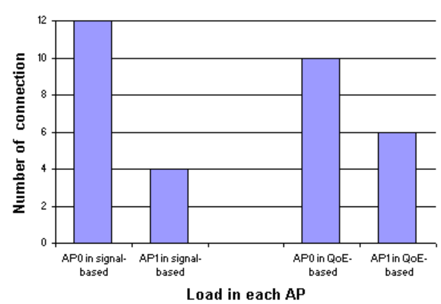
(a) Topologie de réseau



(b) Notes MOS individuelles des nœuds



(c) MOS de l'ensemble du réseau AP0



(d) Répartition de charge entre les APs

Figure 10: Comparaison des performances entre le mécanisme basé sur la QdE et celui basé le signal.

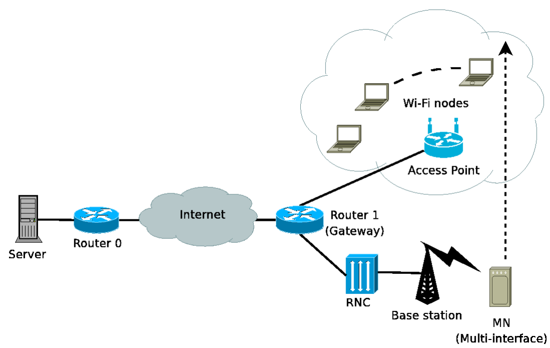
- **Sélection du réseau d'accès - Environnement hétérogène**

Puisque le déploiement du réseau de prochaine génération (4G) commence à se propager à travers le monde, il est difficile pour les utilisateurs de sélectionner le meilleur accès parmi plusieurs technologies existantes. Par conséquent, un autre mécanisme de sélection est présenté [8]. Il étend le précédent en prenant en considération l'expérience utilisateur ainsi que d'autres facteurs comme le coût et la mobilité, cette fois dans un environnement hétérogène.

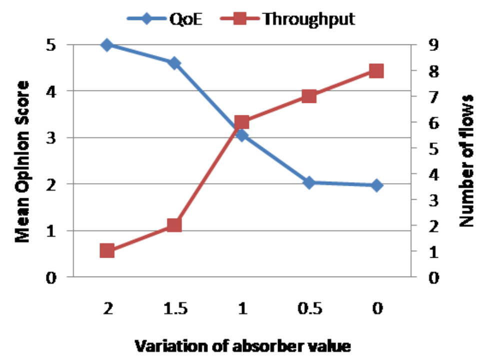
Des moyens de communication utilisés dans ce type de réseau peuvent être les messages de signalisation du standard IEEE 802.21 MIH (Media Independent Handover) [31]. Pour la décision, la fonction objective OF est définie par la somme de chaque critère i (C_i) multiplié par leur poids (w_{ci}). Les poids peuvent être modulés en fonction des exigences du client, et la somme de tous les poids est égale à 100. La valeur de chaque critère est normalisée par sa valeur maximale, ce qui donne une note comprise entre 0 et 100 pour chaque réseau. Une fois le calcul de OF effectué pour chaque réseau candidat, l'utilisateur hiérarchise les différents réseaux et sélectionne celui qui a le meilleur résultat. Si la demande de connexion au premier AP choisi ne peut être satisfaite par l'opérateur, la station essaie le suivant dans la hiérarchie, et ainsi de suite. Pour garantir la qualité de l'application, une marge est ajoutée à la note requise pour absorber la dégradation. La meilleure valeur de marge est choisie selon les résultats dans la figure 11(b).

Le scénario est présenté dans la figure 11(a). Le nœud mobile (MN) est un terminal multi-interfaces équipé d'interfaces WLAN et UMTS. Au début, le seul réseau présent est l'UMTS donc le MN commence sa connexion via ce réseau. Le MN se déplace durant la connexion jusqu'à ce qu'il entre dans la couverture du réseau WLAN. Le MN doit alors décider soit de rester sur le réseau UMTS soit de changer pour aller sur le réseau WLAN. Dans ce scénario, le WLAN est déjà chargé par plusieurs connexions existantes et des nouvelles demandes.

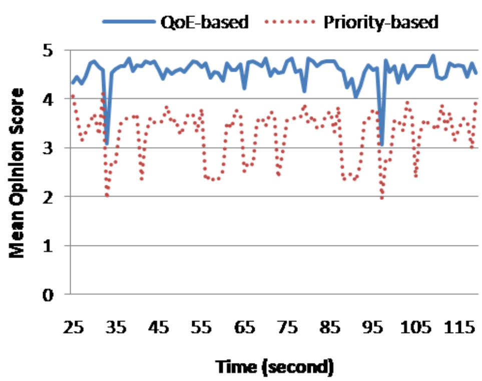
Ce mécanisme est comparé avec une approche basée sur la priorité, actuellement utilisée sur de nombreuses implémentations de Mobile IP sur le marché. Les résultats obtenus montrent que le mécanisme proposé donne de meilleurs résultats lorsque l'on souhaite garantir à la fois la qualité d'expérience du nœud mobile (figure 11(c)) et les utilisateurs en cours dans le réseau ciblé (figure 11(d)). La répartition de la charge est également préférable puisque le réseau UMTS garde le trafic de MN. Ces résultats montrent que même avec un mécanisme simple, nous pouvons déjà observer une amélioration considérable des performances.



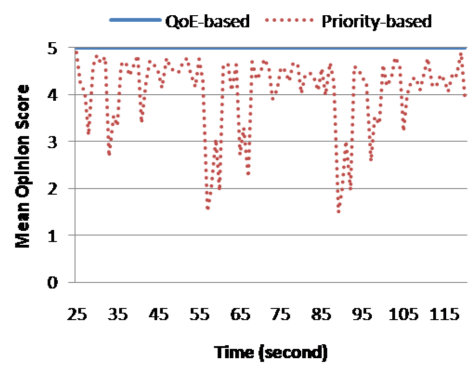
(a) Topologie de réseau



(b) Résultats avec différents marges



(c) MOS global des nœuds dans Wi-Fi



(d) MOS du nœud mobile (MN)

Figure 11: Comparaison des performances entre le mécanisme basé sur la QdE et celui basé sur la priorité.

5. Conclusion et Perspectives

Conclusions générales

Ce document de thèse fournit une étude approfondie de la gestion de ressources en utilisant la qualité d'expérience ou QdE, un nouveau concept de qualité qui a récemment émergé dans les réseaux multimédias d'aujourd'hui. Une méthode d'évaluation appropriée (PSQA) a été choisie afin de mesurer la QdE en temps réel. En utilisant un apprentissage statistique avec un réseau de neurones aléatoires, cette méthode reproduit l'expérience utilisateur en utilisant les informations du trafic réseau en temps réel.

Avec cette mesure automatique de la QdE, de nombreuses orientations de la gestion des ressources ont été explorées. Ceci comprend la gestion à la fois côté réseau et côté utilisateur. Les mécanismes côté réseau sont le contrôle d'admission, l'adaptation de débit et l'ordonnanceur. L'indicateur QdE est utilisé pour tous ces mécanismes. En ce qui concerne le côté utilisateur, la gestion des connexions avec le mécanisme de sélection du réseau a été étudiée. Les investigations ont commencé en milieu homogène et ensuite dans un environnement hétérogène. Les résultats obtenus (satisfaction des clients, utilisation de la bande passante, équilibrage de charge, et équité) illustrent les bonnes performances du déploiement de la QdE et son utilisation comme indicateur dans la gestion des ressources.

Il peut être remarqué que seuls les cas de transmission vidéo ont été étudiés, mais les mêmes idées peuvent être appliquées à d'autres types d'applications multimédias. En outre, comme la QdE est indépendante du contexte, elle peut également être déployée dans d'autres technologies de réseau ou architectures.

QdE dans la gestion des ressources

Cette section traite des limitations et remarques concernant l'utilisation de métriques QdE dans la gestion de réseaux. Comme l'outil PSQA a été déployé pour mesurer la QdE, ses limites et des remarques le concernant seront également discutées.

La QdE devient progressivement un facteur essentiel pour la gestion des ressources. Puisque les réseaux deviennent de plus en plus hétérogènes, des travaux futurs pourront porter sur la gestion des ressources dans un tel environnement en utilisant la QdE (qui est indépendante du contexte) comme métrique. L'hétérogénéité ne concerne pas seulement la technologie des réseaux, mais aussi les applications, les utilisateurs, les appareils, etc. Avec la croissance des applications multimédias dans les réseaux de prochaine génération, divers types de trafics se répandront sur ces réseaux. La différenciation des services sera nécessaire pour traiter tous les types d'applications en fonction de leurs caractéristiques et de leurs exigences. Différents traitements seront nécessaires pour satisfaire l'utilisateur et tout en optimisant l'utilisation des ressources.

Un autre point concerne la qualité d'expérience garantie aux utilisateurs finaux. L'opérateur de réseau doit tenir compte du fait que le service fourni est garanti par un MOS moyen ou minimum en trouvant le meilleur compromis pour l'opérateur et les utilisateurs. Si le service garanti est en termes de note moyenne pendant la durée de la connexion, il est acceptable d'avoir quelques instants de MOS bas et certains autres moments avec un MOS haut pour compenser. De même, si le service est garanti en termes de valeur minimale, l'opérateur de réseau doit s'assurer lors de la connexion que l'utilisateur percevra au moins cette valeur minimale. Il est utile de rappeler que la qualité d'expérience est subjective et, en général, un utilisateur est plus sensible au moment de mauvaise qualité donc l'utilisation d'une valeur minimale peut être plus risqué pour l'opérateur. Dans tous les cas, un SLA approprié doit être établi à l'avance en indiquant les spécifications du service offert et la responsabilité de chaque partie.

En ce qui concerne la mise en œuvre et l'utilisation de PSQA, on remarque que PSQA est un bon outil pour mesurer la qualité d'expérience en temps réel, mais il faut mentionner que même si la sortie de PSQA (qualité d'expérience) est indépendante du contexte, les entrées de PSQA et sa méthodologie sont quant à elles spécifiques au contexte. Avec cette méthodologie, le RNN validé fonctionnera seulement avec la même application et dans le contexte où il a été validé. Par exemple un RNN validé avec une application de streaming vidéo ne sera pas précis lors de son utilisation pour mesurer une application VoIP. En effet ces deux applications ont des caractéristiques différentes, se traduisant notamment par des différences dans les facteurs pris en entrée du RNN. Par exemple, les facteurs liés au temps (par exemple, délai et la gigue) sont essentiels dans la VoIP, mais moins important dans le streaming vidéo car il y a la mise en tampon (buffering) du flux avant la lecture. En ce qui concerne l'environnement, la distribution des pertes sur un réseau sans-fil est différente de celle dans le réseau filaire. Néanmoins, le principal avantage est que la procédure d'entraînement est faite une fois pour toute et qu'ensuite l'outil PSQA peut être utilisé pour mesurer la QdE en temps réel autant de fois que souhaité.

Perspectives

On peut remarquer qu'il y a divers applications sur le réseau d'aujourd'hui, chacune avec ses propres besoins. Un service sur mesure doit être fourni par le réseau opérateur en termes de besoin en bande passante, la sensibilité au délai, etc. Le même argument s'applique aussi aux utilisateurs du réseau. Les utilisateurs privilégiés (payant généralement un prix plus élevé) devraient avoir un meilleur accès aux ressources comparé aux utilisateurs de plus basse priorité. La gestion des ressources doit être consciente de ces facteurs. Un sujet potentiel pourrait donc être la gestion de "différenciation de service" en tenant compte, par exemple, de l'expérience utilisateur, de la priorité de classification des services ou des utilisateurs, etc. Deux applications représentatives, à savoir la vidéo et la voix sur IP, pourraient être envisagée ainsi que le trafic background. La gestion sera basée sur l'expérience utilisateur afin d'être plus souple et plus efficace que celle fondée sur des paramètres techniques. Par exemple, des mécanismes d'ordonnancement améliorés pourraient être proposés pour fournir une qualité appropriée pour chaque application et chaque utilisateur.

En outre, il serait utile que nous puissions prédire l'expérience utilisateur (prédiction du MOS). Quelques travaux ont déjà commencé sur ces aspects. Cela peut être fait grâce à l'apprentissage, le mappage, ou d'autres stratégies de modélisation. Si la prédiction précise de la QdE est disponible, on peut imaginer tout un système de réseau informatique basé sur la QdE pour la gestion des ressources. Par conséquent, il serait intéressant d'étudier la possibilité et la faisabilité de concevoir une telle architecture. De nombreuses questions doivent être examinées : les entités de contrôle des ressources du réseau, les communications entre les entités du réseau, la facturation, les questions de sécurité, etc. En outre, l'hétérogénéité peut également concerner d'autres éléments que la technologie du réseau. Et la question d'interopérabilité va devenir cruciale et devra être étudiée afin de rendre toutes ces hétérogénéités compatibles tant au niveau des technologies des réseaux sans-fil que des mécanismes de gestion des ressources.

Outre les aspects d'hétérogénéité, les recherches pourront continuer sur d'autres architectures telles que les "réseaux overlay" comme par exemple les réseaux pair-à-pair, ou même les réseaux CDN (content delivery network) qui émergent. Avec ces architectures de réseau, il sera avantageux d'étudier comment la gestion des ressources peut être améliorée en utilisant l'indicateur QdE. Par ailleurs, comme dans le présent document la gestion des ressources est étudiée du côté réseau et du côté utilisateur; il serait également intéressant d'étudier la gestion des ressources de bout en bout.

Chapter 1

Thesis Introduction

1.1 Problem statement

At the beginning of networking era, connections are established via cables or what we call *wired network*. This type of network provides high bandwidth and stable condition, making it easier to manage network resources. With progress on network technologies, wireless and mobile networks are increasingly emerged as users want to be connected anywhere and anyhow. Moreover, the ability to connect user to the network using air interface facilitates connection establishment greatly; as a result, wireless networks and users are now everywhere. Many devices and applications are released to be operated on this type of network. A personal computer (PC) today can work on both wired and wireless environment; more specifically, mobile devices can now connect user to the Internet via different access networks/technologies simultaneously. Meanwhile, users are more and more interested in multimedia applications as we currently observe tremendous growth of this traffic on the network. In addition, network users also become more experienced and traditional ways of measuring quality using technical network parameters do not accurately reveal quality perceived at user. Therefore, it is now more interesting to measure the quality in terms of user perception of the provided service or what we call *quality of experience (QoE)*.

Obviously, the two principal actors in this context are *network operator* and *network user*. The role of network operator is to provide services to users via different access networks and technologies; network user is then client of provided services. It can be noticed that in this business model, users play an important role as their satisfaction is fundamental to operator's benefit. The key resource that needs to be managed is the *bandwidth*, which is restricted and varying because of wireless nature. For network operator's perspective, bandwidth needs to be distributed efficiently in order to yield the most revenue. For network users, they want to select the best network, which will provide the best quality with the lowest price. In such situation, resource management is crucial since efficient mechanisms can help satisfying both parties.

There exist many ways of managing network resources. At network side, operator can deploy admission control mechanism that manages incoming traffic by filtering (admit/refuse) new connection in order to control amount of traffic in the network. Adaptation can also been done at point of attachment such as access point or base station; for example, transmission rate can be adjusted in order to better suit the current condition of the network. The packets can be scheduled according to quality and/or priority of users or traffic class as well. As for user side, network users can manage their connections using mechanism like network selection to help them choosing the best network among several accessible today. Managing network resource in this context is a complicated task due to different factors; descriptions of the points that are going to be considered in this dissertation are listed below:

- First of all, the *wireless nature* of the network makes management becomes more difficult. Due to its open environment, wireless network is prone to all types of interference and disturbance. As a result, network condition varies often; hence, guaranteeing service quality can become a complex issue.

- Second factor is the *increasing amount of traffic* due to rising number of Internet users. Many progresses have been done and user terminals are now affordable by almost everybody, network connections are various and accessible everywhere making it much easier for people to get a connection. This phenomenon increases difficulty for managing resources since increased traffic results in higher congestion and also more interferences in wireless environment.

- Another important factor is the *rise of multimedia applications* in wireless networks. With this type of application, managing resource using technical parameters is no longer appropriate as it is too conservative. In such approach, limits are fixed for technical parameters and operator has to manage resource accordingly. Since many multimedia applications generate variable bit rate traffic, handling quality using, for example, bandwidth restriction is not enough, especially in wireless environment where network resource is scarce and radio condition changes often.

- Resulting from the growth of multimedia applications, Quality of Service (QoS) becomes less significant and the notion of *Quality of Experience (QoE)*, or sometimes called user experience[1], is becoming more meaningful to network user. QoE reveals the quality of a service as perceived by user. As the final objective of every service is user satisfaction, quality of experience is thus the most important concern. Therefore, network operators who wish to maximize their profit by optimizing resource utilization also have to keep user fidelity that results directly from user satisfaction.

[1]User experience and quality of experience in this document have the same meaning and they will be used interchangeably from now on.

- *Variety of applications* in wireless networks today also makes resource management very hard to deal with. Different types of application (VoIP, video streaming, interactive games, emails, FTP, etc.) have different requirements in terms of bandwidth, delay, jitter, etc. Hence, appropriate and differentiated treatment is needed for each type of applications if we want to satisfy user expectation.
- Finally, *heterogeneity in access network* or what we could call heterogeneous network environment becomes reality. As today user's devices are equipped with several interfaces enabling the connection to different network technologies (Ethernet, Wi-Fi, Cellular, Satellite, etc.), even in simultaneous manner. Diverse technologies have their diverse characteristics and they can be combined together in order to provide a heterogeneous system, a very powerful system enabling all classes of applications to find the right access network. Hence, the arrival of this type of environment needs special treatments and increases complexity of management problem dramatically.

To briefly summarize, this dissertation will focus on resource management problems in wireless networks. The topics that will be handled concern bandwidth and connection management. Business aspect such as pricing and SLA (Service Level Agreement) are out of the scope of this dissertation. In case of heterogeneous environment, one network operator possessing different network technologies is assumed. Additionally, security aspects such as authentication and authorization are not the focus of this dissertation neither. Therefore, a server, type AAA (Authentication, Authorization, and Accounting) server, is assumed to be present in the network and it is the entity that manages all these aspects efficiently.

1.2 Motivations and objectives

Resource management in wireless networks can be handled regarding to different points of view. In terms of *technology-oriented*, each wireless technology can be managed independently and solutions can be constructed for each of them separately. Or, in terms of *environment-oriented*, issues can also be classified according to environment type (e.g. homogeneous or heterogeneous) and solutions can be established for each distinct category of network. Furthermore, in terms of *layer-oriented*, management can also be done at different layers of OSI[2]. Many researchers have tried to manage the network in the IP, MAC, or PHY layers separately or some time collaboratively as in cross-layer design, for example.

Other than using previous classifications, the work in this document will be classified according to the two main actors, namely network and user. The management

[2] The Open System Interconnection Reference Model (OSI Reference Model or OSI Model) is an abstract description for layered communications and computer network protocol design. It was developed as part of the OSI initiative [32].

is then categorized into two approaches: *Network-centric* and *User-centric* regarding network and user perspectives respectively. Several solutions will be proposed. From network perspective, operator should be able to guarantee a certain level of quality in order to obtain client's fidelity and thus good revenue, even in wireless and varying network condition. As for user perspective, user should be able to select the best available network.

For network-centric approach, final decision is done at network side and network operator's benefit is the principal concern. Even though in some cases user information is also collected for making the decision, the final decision is made at the network operator and management mechanisms are applied in order to reach operator's objective. Network-centric mechanisms are, for example, *admission control* mechanism that filters incoming connections in order to control congestion in the network, *packet scheduler* that can be used to schedule user at the specific time according to network conditions, etc.

On the contrary, for user-centric approach, final decision is held at user terminal and user's benefit is the principal concerns even though some mechanisms may also take into account information from network environment. It can be seen that user does not have much control other than actions concerning user terminal; usually, user-centric mechanisms are related to *network selection* scheme that helps user in choosing the best network among several candidates.

Regarding multimedia applications, it is important to consider not only technical network parameters but also user experience of the provided service. Today's users become more experienced and their expectation is in terms of satisfaction and not in terms of guaranteed network parameters. For that, quality of experience concept should be investigated as it is suitable to network evolution nowadays. We can observe that even though many mechanisms have been proposed in the literature, very few takes QoE into consideration. Since there is a lack in studying the impact of user experience on network management, the goal of this document will be to explore management issues with this new concept of quality. As quality of experience is independent of network technologies and applications, it is thus flexible and it can match perfectly with heterogeneity in network today.

By definition, user experience is related to application layer; however, it can be handled at other layers as well. For application layer, adaptations can be done at end-user or end-server in order to improve quality of the stream. This includes technique like *stream switching* or new codec like *scalable video coding* in which multimedia server can adapt encoding rate dynamically according to network condition. For network layer, quality can be improved if we can control properly the network status. For media access control and physical layer as well, the transmission rate can be adapted in order to suit the physical condition. In this dissertation, the focus will be principally on network and MAC layer where controls can be executed by network operator. Moreover, investigation is also done from the user perspective.

1.3 Thesis contributions

Topics concerning problems stated previously have been investigated. First of all, state of the art in resource management is studied. According to the literature, multimedia application is the problematic issue. This type of application has restricted requirements and it is difficult to guarantee a level of service quality. Many management schemes are proposed but very few of them are interested in user experience. As mentioned earlier that final objective of a service is user satisfaction and thus quality of experience is the most important factor. Meanwhile, network operators should also be satisfied of their profit by optimizing resource utilization. According to that, management mechanisms that aim to satisfy user experience and at the same time to optimize resource utilization are studied in this thesis. For network side, QoE-oriented mechanism such as admission control, rate adaptation, and scheduling have been proposed and network selection mechanisms for user side. Most of them take into account information from both user and network to cover all criteria. The studies have been conducted in different wireless technologies (IEEE 802.11 and Cellular Network) in both homogeneous and heterogeneous way. The obtained results demonstrate that it is feasible and beneficial to use quality of experience as metric to improve network management in the future.

The work presented in this document has been published in the following articles:

- [1] K. Piamrat, A. Ksentini, J.-M. Bonnin, and C. Viho. "Radio resource management in emerging heterogeneous wireless networks". Computer Communications, In Press, Corrected Proof, Available Online, Feb. 2010.

- [2] K. Piamrat, K. Singh A. Ksentini, C. Viho, and J.-M. Bonnin. "QoE-aware scheduling for video-streaming in High Speed Downlink Packet Access", IEEE Wireless Communications & Networking Conference (WCNC 2010), 18-21 Apr. 2010.

- [3] K. Piamrat, A. Ksentini, J.-M. Bonnin, and C. Viho. "Q-DRAM: QoE-based dynamic rate adaptation mechanism for multicast in wireless networks". In IEEE Global Telecommunications Conference (GLOBECOM 2009), pages 1-6, 30 Nov. - 4 Dec. 2009.

- [4] K. Piamrat, A. Ksentini, J.-M. Bonnin, and C. Viho. "Rate Adaptation mechanism for Multimedia Multicasting in Wireless Networks", Sixth International Conference on Broadband Communications, Networks, and Systems (Broadnets 09), pages 1-7, Sep. 2009.

- [5] K. Piamrat, C. Viho., J.-M. Bonnin, and A. Ksentini. "Quality of Experience Measurements for Video Streaming over Wireless Networks", In Sixth International Conference on Information Technology: New Generations (ITNG 09), pages 1184 -1189, April 2009.

- [6] K. Piamrat, A. Ksentini, C. Viho, and J.-M. Bonnin. "QoE-Aware Admission Control for Multimedia Applications in IEEE 802.11 Wireless Networks". In IEEE 68th Vehicular Technology Conference (VTC 2008-Fall), pages 1-5, Sep. 2008.

- [7] K. Piamrat, A. Ksentini, C. Viho, and J.-M. Bonnin. "QoE-based network selection for multimedia users in IEEE 802.11 wireless networks". In 33rd IEEE Conference on Local Computer Networks (LCN 2008), pages 388-394, Oct. 2008.

- [8] K. Piamrat, C. Viho, A. Ksentini, and J.-M. Bonnin. QoE-aware Network Selection in Wireless Heterogeneous Networks. In Research Report RR-7282, INRIA, 2010.

- [33] K. Piamrat, C. Viho, A. Ksentini, and J.-M. Bonnin. Rate Adaptation Mechanisms for Multimedia Multicasting in Wireless IEEE 802.11 Networks. In Research report, IRISA, 2009.

- [34] K. Piamrat, C. Viho, A. Ksentini, and J.-M. Bonnin. Resource Management in Mobile Heterogeneous Networks: State of the Art and Challenges. In Research Report RR-6459, INRIA, 2008.

1.4 Thesis outline

This section provides the outline of the dissertation by giving brief details on the different chapters that present the contribution of the study. Fig. 1.1 illustrates underlying themes from introduction through conclusions and perspectives.

– *Chapter 1: Introduction*
The document begins with this introduction chapter giving description on research topic and problem statement. Then motivations and objectives are presented as well as thesis contributions. The chapter ended with thesis outline, which is explained in more details in the following.

- **PART I: Quality-oriented Resource Management**
  This part provides state of the art and backgrounds on quality-aware resource management topic. It contains two chapters: State of the Art (chapter 2) and Quality of Experience in Resource Management (chapter 3) described as follow:

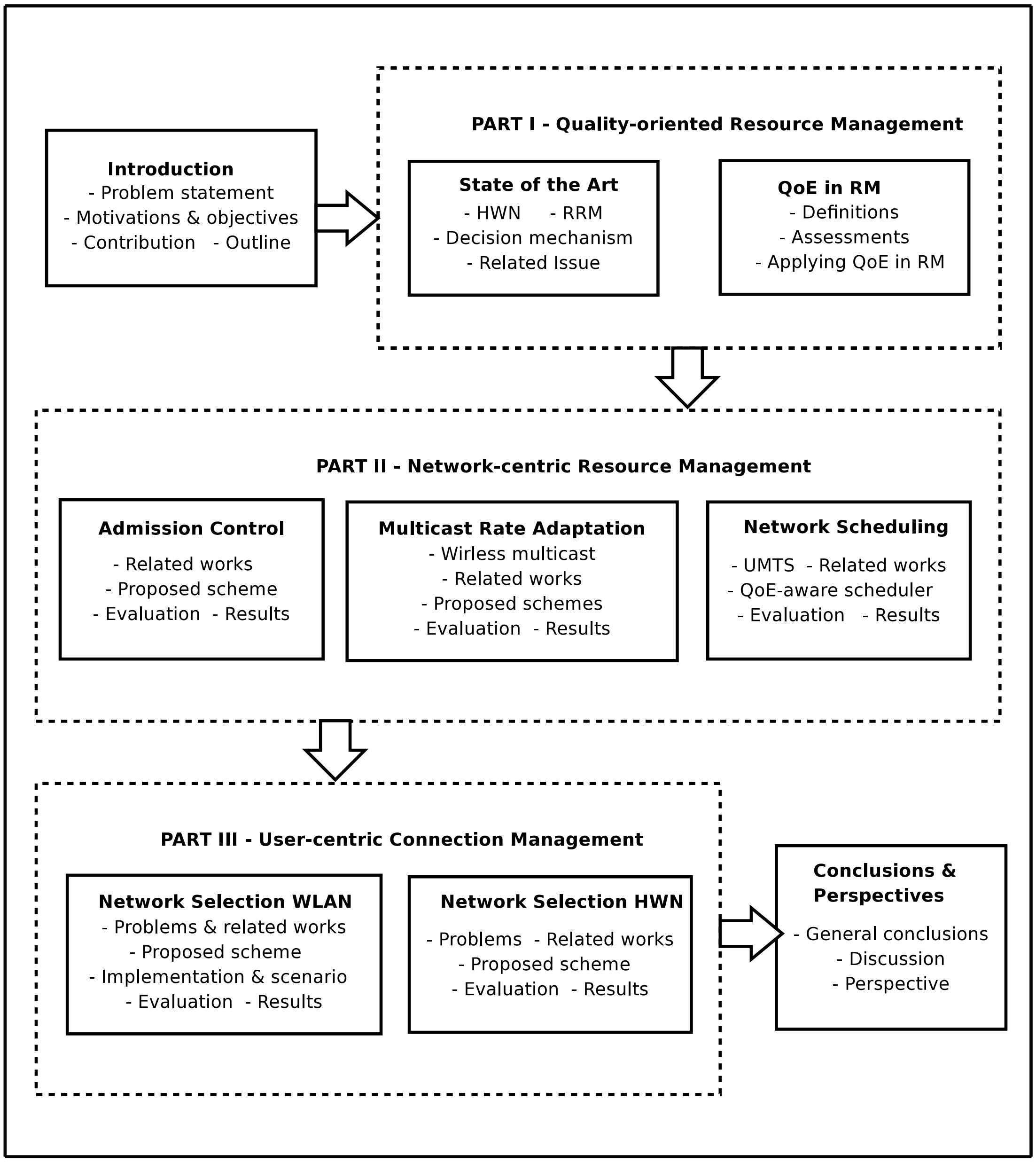


Figure 1.1: Thesis Outline.

  - *Chapter 2: State of the Art*
    This chapter provides state of the art in resource management under heterogeneous wireless network (*HWN*). Backgrounds and characteristics of heterogeneous wireless network are described. Typical management procedure in radio resource management (*RRM*) is explained and discussed. Recent and representative mechanisms in decision making are surveyed and important concerns, such as QoS, mobility, and architecture, are discussed.
  - *Chapter 3: Quality of Experience in Resource Management*
    In this chapter, the concept of quality of experience is introduced along with assessment approaches and their evaluation. The focus is on a technique called Pseudo-Subjective Quality Assessment (PSQA), which enables accurate QoE assessment in real time. After that, research directions on how to use QoE in resource management (RM) are given.

  Management mechanisms are classified into two approaches: *network-centric* and *user-centric*, resulting in the two following parts.

- **PART II: Network-centric Resource Management**
  This part presents network-centric mechanisms proposed for resource management using Quality of Experience as metric. This includes admission control (chapter 4) and multicast rate adaptation (chapter 5) in IEEE 802.11 standard, and then investigation continues on packet scheduling (chapter 6) in one of Cellular network standard called UMTS (Universal Mobile Telecommunications System).

  - *Chapter 4: Admission Control*
    This chapter presents an important problem for network operator, called *congestion control*, in wireless network nowadays. Admission control mechanism is one solution to solve this problem. Related works concerning admission control in this environment are discussed then the QoE-based mechanism is proposed. It provides a solution being aware of user experience. Access point functionality in the scheme as well as interaction between access point and PSQA are also explained. After that, implementation and performance evaluation in network simulator NS-2 are given.
  - *Chapter 5: Multicast Rate Adaptation*
    This chapter begins with introduction to wireless multimedia multicasting including its advantages and drawbacks. Related works concerning rate adaptation mechanism in unicast and multicast environment are provided and discussed. Then two schemes are presented, one in static approach and the other in dynamic approach, to solve the problem whilst being aware of user experience of multicast clients. Then, performance evaluations are given and results are discussed.

– *Chapter 6: Packet Scheduling*
This chapter presents the study of another popular technology, UMTS, one of the recent cellular networks. Related works concerning network scheduling are described and QoE-aware schedulers are presented along with performance evaluation and results.

- **PART III: User-centric Connection Management**
This part investigates management problem from the user perspective. It presents user-centric mechanisms such as network and handover selection mechanism in homogeneous environment (chapter 7) and heterogeneous environment (chapter 8) respectively.

  – *Chapter 7: Network Selection in Wireless Local Area Networks*
  This chapter describes current problem of network selection and existing solutions. Then, network selection mechanism based on quality of experience is presented together with access point/mobile host functionalities and interactions with PSQA. The chapter ends with implementations and performance evaluations.

  – *Chapter 8: Network Selection in Heterogeneous Wireless Networks*
  This chapter investigates network and handover selection problem, this time, in heterogeneous environment. The problem statement and related works are given along with the proposed mechanism. Then performance evaluation is conducted and obtained results are discussed.

– *Chapter 9: Conclusions and Perspectives*
Finally, the document ends with this chapter providing conclusions and perspectives. Different discussions and conclusions of QoE-aware resource management in wireless networks are provided. Furthermore, open research directions are also considered.

Part I

Quality-oriented Resource Management

The Part I consists of two chapters with the objective to provide readers a better understanding and good backgrounds on quality-oriented resource management. Definition of environment and analysis of resource management procedures will be described. State of the art in resource management based on quality will be given along with detailed investigations of current advances. Since research in this topic has been extensively studied in recent years and many schemes have been proposed, a survey of representative approaches will be given. Techniques deployed for decision mechanisms are described and classified into three categories: network-centric, user-centric, and collaborative. Moreover, discussions on QoS and mobility supports as well as architectural design and media adaptation are also included. Besides, as multimedia applications have emerged drastically, representing quality of a provided service using technical or QoS parameters is no longer suitable. Therefore, reader will be introduced to a new concept of quality called user experience or quality of experience. Definition and fundamental elements will be described and an appropriate measuring tool will be selected. Then discussion of how to deploy this concept in real-time resource management will be discussed along with examples of use case.

Chapter 2

State of the Art

2.1 Introduction

Deployment of heterogeneous wireless networks is spreading throughout the world as users want to be connected anytime, anywhere, and anyhow. Meanwhile, these users are increasingly interested in multimedia applications, which require strict QoS support, such as video streaming and Voice over IP (VoIP). Provisioning network resources with such constraints is a challenging task. In fact, considering the availability of various access technologies (Wi-Fi, WiMAX, or Cellular networks), it is difficult for a network operator to find reliable criteria to select the best network that ensures user satisfaction while maximizing network utilization. Designing an efficient management mechanism, in this type of environment, is mandatory for solving such problems.

In order to have a good understanding on the topic, this chapter provides comprehensive survey on state of the art in quality-aware resource management. The chapter is organized as follow. We begin with definition and description of *heterogeneous wireless networks* in Section 2.2, and then a thorough analysis of *resource management* procedures and their interactions are provided in Section 2.3. A review of recent advances in decision mechanism is presented in Section 2.4. A classification of these works according to who is making management decisions is proposed; that is, the decision making is based on: *network-centric*, *user-centric*, or *collaborative* approach between network (operator) and users. Moreover, since decision making alone may not be sufficient to guarantee an efficient management, Section 2.5 also gives an overview of related topics such as: *QoS support*, *mobility support*, *architectural design*, and *media adaptation*. Finally, Section 2.6 draws conclusions.

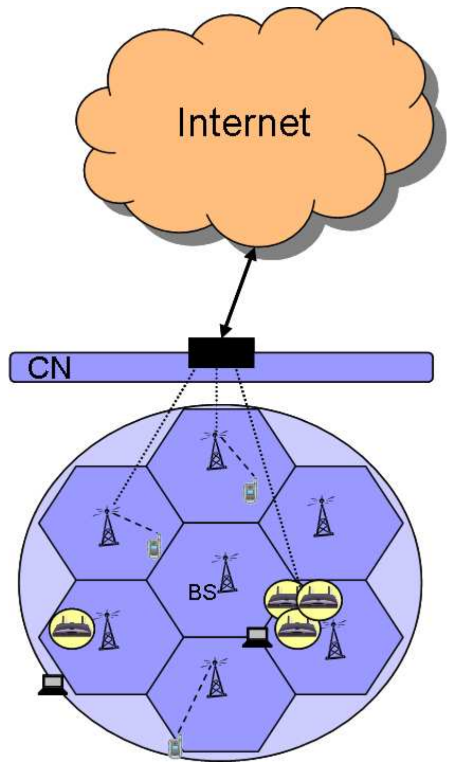


Figure 2.1: Heterogeneous Wireless Network (HWN) can be described as a set constituted of several wireless technologies, which connect users to the Internet through a *core network*, also known as *backbone network*. The wireless technologies involved in HWNs are called *access networks*, and their coverage can overlap to cover hot areas (hotspots).

2.2 Heterogeneous Wireless Networks

The evolution of network technology has led to a deployment of various access networks such as Cellular Networks (*GSM*, *UMTS*, *HSPA*, *LTE*), Wireless Local Area Network (*IEEE 802.11 family*), Digital Video Broadcasting (*DVB-T*, *DVB-S*, *DVB-RC*, *DVB-H*), or Broadband Wireless Communication (*IEEE 802.16 WiMAX family*). In this section, the focus will be on wireless technologies, which are wide spreading nowadays. In fact, a heterogeneous wireless network (HWN) is composed of two or more wireless access technologies, empowered by their overlapping coverage. Fig.2.1 illustrates how users are connected to the Internet through access networks and core network (CN). Each access technology involved in HWN has its own characteristics in terms of coverage, QoS support, and operational costs. Fig.2.2 illustrates different sizes of coverage provided by these access technologies. Examples of their characteristics, in terms of bandwidth, coverage, cost, and application, are presented in Table 2.1.

The arrival of HWN brings out important advantages. Since users are now equipped with multi-interface terminals, they can get connectivity from different wireless technologies. Thus, an attractive property of HWN is the ability to provide the best features of each individual network. One example could be the coexistence (overlapping) of 3G cellular network and Wireless Local Area Network (WLAN). Cellular networks

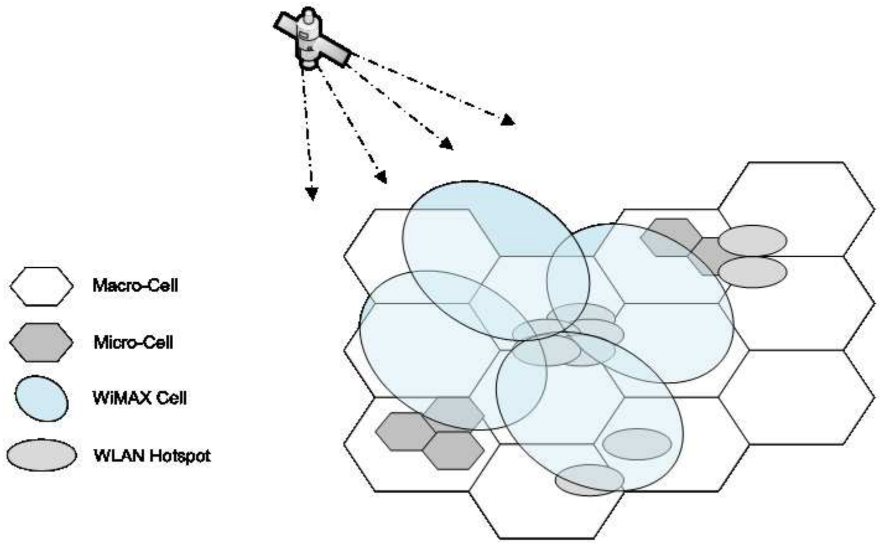


Figure 2.2: Heterogeneous Wireless Network.

Table 2.1: Wireless Technologies.

Class	**Technology**	**Data Rate**	**Range**	**Cost**	**Application**
Cellular	2G: CDMA, GSM 2.5G: GPRS, EDGE 3G: UMTS 3.5G: HSDPA 4G: LTE	$\leq$ 20 Kb/s 30-90 Kb/s 2 Mbps 0.384-14.4 Mbps $\geq$ 100 Mbps	Cellular network	Monthly charge	Cellular phone, multimedia applications, or SMS/MMS
WLAN	ZigBee	0.02-0.2 Mbps	70-300m	Free	Sensor network
	802.11a 802.11b 802.11g 802.11n	54 Mbps 11 Mbps 54 Mbps 100 Mbps	100 m 100 m 100 m 100 m	Free	LAN Internet
WPAN	802.15 Blue tooth	0.8-1 Mbps	$\leq$10 m	Free	Cable replacement
	Ultra-Wideband	50-100 Mbps	10-30 m	Free	Synchronization and transmission of video/audio
WMAN	802.16 WiMAX	70 Mbps	50 km	Free	Metropolitan area broadband Internet

such as UMTS or HSPDA support low bandwidth over a wide geographical area while WLAN, based on IEEE 802.11g or the upcoming IEEE 802.11n, can provide relatively high bandwidth (up to theoretical 300 Mbps) in a smaller coverage. All together they can provide wider ranges of service and quality than in homogeneous environment. Multi-mode users can connect to the best network and profit from the best QoS offered by the heterogeneous system.

On the other hand, for network operators, an important motivation in deploying a heterogeneous system is higher revenues through exploiting the complementary advantages of each access technology. This is obvious if the heterogeneous system belongs to one network operator. Otherwise, different operators will need to collaborate and agreements would have to be established defining responsibilities of each party. In this chapter, the focus will be on how resources in such an environment can be efficiently managed. As mentioned earlier, business aspects like pricing or security among network operators or between network operators and users are not in the scope of this dissertation and this chapter, except when it is explicitly specified otherwise.

It is worth mentioning that the implementation of HWN is expanding today (year 2010) as supporting devices have been introduced to the market, for example, Apple's iPhone, BlackBerry, and Nokia N Series, that enable users to connect to at least four radio interfaces, including GSM, 3G, WLAN and Bluetooth or DVB-H in the near future. Nowadays, users are already able to initiate connection through any of these technologies simultaneously.

However, designing an efficient Radio Resource Management (RRM) framework in the context of HWN is not simple. RRM concerns overseeing the distribution of radio resources to different users, or different classes of users, in order to maximize the number of services delivered (and thus network operator's revenues) while ensuring user satisfaction. This is a difficult task as there is tradeoff between user satisfaction and network resource utilization. Typical characteristics of HWN challenge traditional arguments for designing management frameworks. Managing the resources of an access technology in an HWN independently of other networks, in which it is overlaid, risks underutilization and resource mismanagement. To obtain an efficient framework, network operator has to consider different procedures and functionalities with the assistance of users' terminal.

2.3 Radio Resource Management (RRM)

To the best of my knowledge, there is not any recent survey of resource management in HWN especially on decision mechanism. Related works are a comparison of four IST (Information Society Technology) architectures in [35], a discussion on IST projects in [36] (both mainly focus of architectural aspect), and a survey on common radio resource management in [37] (only focuses on a combining system of Cellular and

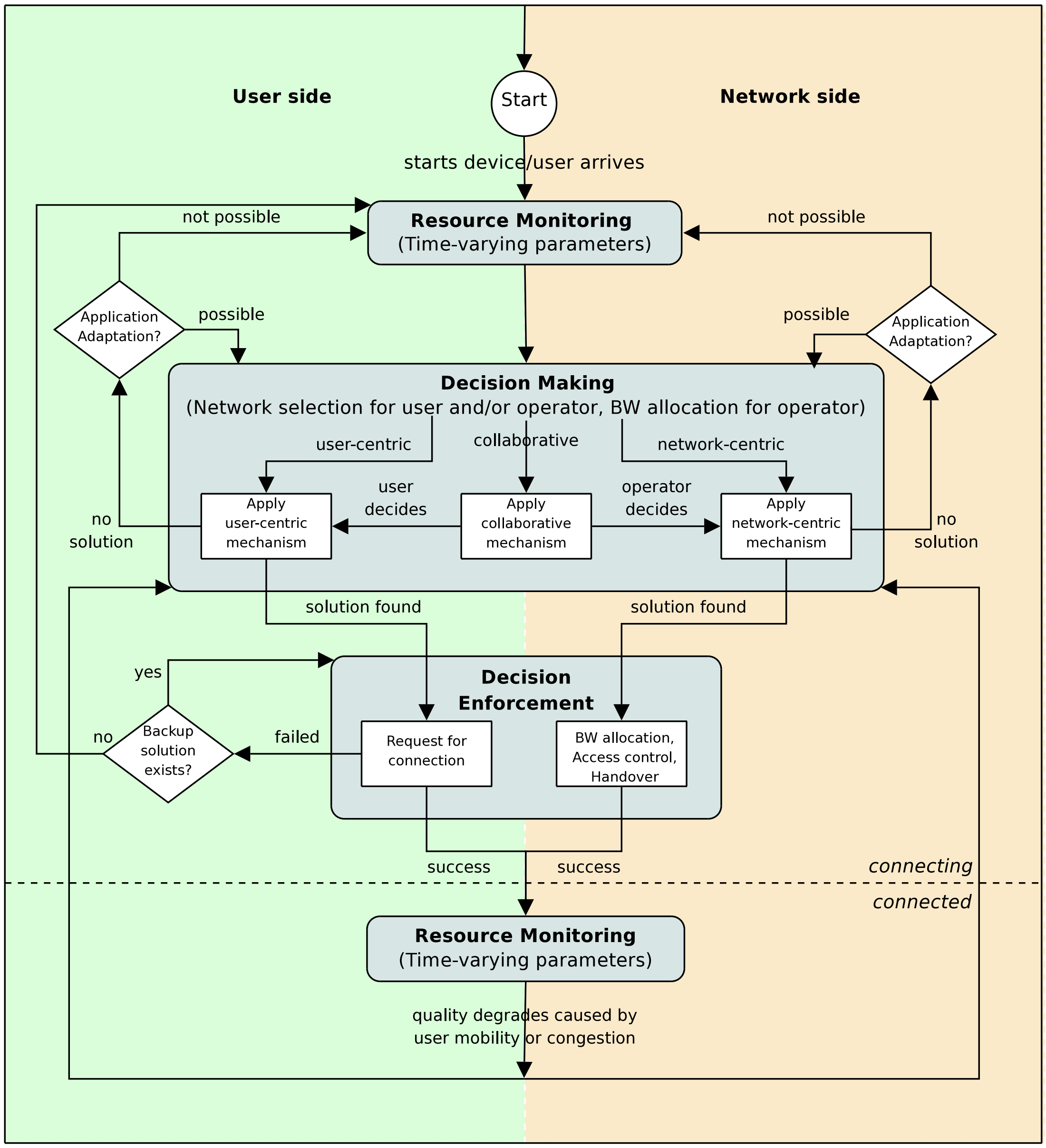


Figure 2.3: A Global Vision of Resource Management in HWN.

Wi-Fi). With increasing number of techniques deployed in radio resource management nowadays, it is interesting to carry out an investigation. Therefore, in this chapter, a comprehensive survey on resource management in HWN will be presented. The goal is to provide a better understanding of resource management in this type of environment.

Generally speaking, efficient management framework for HWN involves one or all of the following procedures and functionalities: (i) *Resources Monitoring*; (ii) *Decision Making*; (iii) *Decision Enforcement*. Fig. 2.3 depicts a global vision of resource management in HWN. It shows the interaction of the procedures when considering users' connection process. It can be seen that these procedures are complementary and they are related to each other; for example, decision making is mainly based on resource monitoring and decision enforcement is performed after decision making. Details of these procedures and their interactions are given in the following.

1. **Resources Monitoring -** It is the phase in which information is gathered; this data comes from users, networks, or both. Collecting information can vary from one decision mechanism to another; it will be considered as input for making decisions. We can see from the Fig. 2.3 that resource monitoring is situated at two different places: before connecting to the network and after the connection establishment. The first resource monitoring is aimed to collect information for first-time connection (*Network Selection* or *Bandwidth Allocation*); if there is no solution meaning that existing networks do not correspond to the requirements, user may have to modify his/her requirement (*Application Adaptation*) in order to find the appropriate network. If adaptation is not possible then user will have to wait for a better condition by returning to monitor resource again. The second monitoring phase is aimed to observe the ongoing connection state, it is used to trigger network adaptation when undesirable event happens; for example, user moves out of current cell (mobility) or network congestion. In these cases, the decision has to be made again considering the current condition.

   Referring to its nature, the information used for making decision can be separated into two categories: *Pre-Determined* and *Time-Varying* factors as listed in Table 2.2. Factors in the former category are pre-defined and remain unchanged for a certain period of time whereas the ones in the latter change in time. Pre-determined factors are taken into consideration as initial policy or preference; they also include constraints of application and capabilities of technologies and equipments. On the other hand, time-varying factors are monitored continuously; they are mainly network quality parameters.

Pre-Determined	**Users preference**: cost, security, power, visual quality, etc. **Providers preference**: cost, trust, security, load balancing, dropping and blocking probabilities, user priority, topology, etc. **Application constraints**: QoS constraints, application context, application requirements, adaptation ability, minimum required bandwidth, maximum loss rate, latency allowed, delay bounds, traffic specification, etc. **Capabilities**: network capability, network equipments capability, access technologies capability, access point bandwidth and queue, up/downlink bandwidth, modulation scheme, terminal capability: CPU, memory size, display I/O, transmitted power, battery, network interface, built-in application, software platform, etc.
Time-Varying	**Availability**: network load, available radio coverage, visible AP, maximum saturation throughput of AP, transmission bandwidth, cell diameter, bandwidth per user, traffic intensity/connection arrival process, connection holding time, average number of connection, bandwidth utilization, data rate, user activity history, available service, variety of services, etc. **Radio-related**: SINR (signal to interference plus noise ratio), SNR (signal to noise ratio), RSS (received signal strength), SIR (signal to interference ratio), SER (symbol error rate), PSNR (peak signal to noise ratio), radio condition (path loss), CIR (carrier to interference ratio), etc. **Quality-related**: BER(bit error rate), MSE(mean square error), handover latency, loss, dropping rate, delay, jitter, throughput, response time, burst error, etc.

Table 2.2: Pre-determined and Time-varying Factors.

2. **Decision Making** - It is the phase in which decisions are made. Most of time, these decisions are made at the network operator for which is called it network-centric approach, however they can also be made at user terminals (user-centric approach), or some time decisions are made by the collaboration between both sides (collaborative approach). Two main decisions to be made are *Bandwidth Allocation* (how to allocate bandwidth from different networks to users) and *Network Selection* (how to select the best available network for a connection). In HWN, bandwidth allocation can also mean the distribution of bandwidth from several network technologies allocating to one connection; in this case, the resource is called *Joint Resource*. It can be seen from Figure 2.3 that, in collaborative approach, final decision will be made only by one of the two actors (either user or network operator), and then the following steps will correspond to either user-centric or network-centric approach. Decision making represents the heart of RRM; therefore, Section 2.4 will discuss this topic in more details.

3. **Decision Enforcement** - It is the phase in which decisions are enforced/executed. In user-centric approach, this phase is done by ensuring that connection request to a selected network is successful; if it is not, user will try backup solution until the last one. If there is no more solution to try and user still cannot get connection, it will have to go back to monitoring step and wait for new condition. This situation can occur when network refuses incoming request in order to protect overall performance. In network-centric approaches, network selection is enforced using admission control mechanism to filter or direct (guide) access to networks according to the decision made in previous step. Moreover, the decision to move users to another network within the same or to different technology is executed by mechanisms such as vertical and horizontal handover respectively. For bandwidth allocation, the operator distributes bandwidth according to the decision made. We can notice that in user-centric approach, the obtained solution(s) are not always achieved if the network does not accept the request; contrary to network-centric approach where solutions are always achieved since it is the network operator who controls all the resources (HWN owned by a single operator).

2.4 Decision Mechanisms

Defining efficient RRM framework, particularly the decision mechanism, for HWN has attracted many research activities where different solutions have been introduced. In this section, a survey of the most recent and representative schemes, dealing with resource management problems under HWN environment, is given. According to who is benefiting from the decision, solutions are classified into three approaches: *network-centric*, *user-centric*, and *collaborative* approaches as presented in Fig. 2.4.

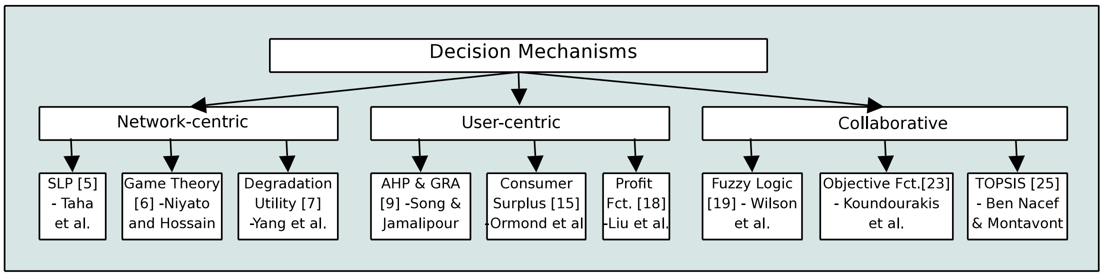


Figure 2.4: Approaches in Decision Making.

2.4.1 Network-centric Approach

In this approach, decisions are made at network side and they are based mainly on the network operator's profit even though some mechanisms may take into consideration user's requirements before making decision. Schemes in this approach deal with how network can optimize its bandwidth and thus bandwidth allocation problem is the important concern. In this subsection, recent techniques and their representative schemes are presented.

- **Stochastic Programming**

  Stochastic programming (SP) [38] is a mathematical technique, which is used in decision making under uncertainty. In [39], the authors deploy SP to design a proactive allocation mechanism. The scheme actually uses a subset of SP called *stochastic linear programming (SLP)* to handle probabilistic nature of demands in HWN. In the exemplary scenario, a single data service of fixed bandwidth requirement is provided by cellular network and WLAN. The idea is to associate probabilistic *demands* with predetermined significant probabilities, then formulate given scenario with *allocation*, *underutilization*, and *rejection* along with the predetermined probability. The goal is to obtain maximum allocation in both networks while minimizing cost of underutilization and demand rejection.

  Here is the formulation for Single Common Service with Probabilistic Demands (SCS-PD). Let S be the set of all possible scenarios. In every scenario $s \in S$, the demand $D_{ij}(s)$ takes on specific values with a predetermined probability $p_{ij}(s)$. The index ij is used to distinguish between entities related to different types of users. Entities indexed with $i = j$ are related with users to be admitted into one network, while entities indexed with $i \neq j$ are related with users changing networks. $D_{ij}(s), R_{ij}(s)$ and $A_{ij}(s)$ respectively refer to the demand, rejection (unsatisfied demand) and allocation for users ij, such that $R_{ij}(s) = D_{ij}(s) - A_{ij}(s)$. The demand uncertainty can be imposed on Program SCS-DD through the allocation-rejection-demand constraints, where the penalty can be applied to the rejection. In this manner, let profit per allocated ij user be x_{ij}, the costs of unit underutilization (U) for network j and the interconnection be y_i and y_v respectively. The penalty (cost) of unit rejection is $z_{ij}(s)$. As such, the return function to be maximized becomes

$$\prod_{SCS-PD} = \sum_{\forall i,j} x_{ij} \times A_{ij} - \sum_{c=\{j,v\}} y_c \times U_c - \sum_{\forall i,j} \sum_{s \in S} p_{ij}(s) \times z_{ij}(s) \times R_{ij}(s)$$

Discussion: To the best of my knowledge, this scheme is the first mathematical attempt that addresses joint resource management, in which user bandwidth is provided by several access networks in the HWN. However, the scheme is designed for supporting single common service with fixed required bandwidth, which is not appropriate to variety of services along with various bandwidth requirements in networking today. Moreover, to our knowledge no future work has been conducted for more realistic or more complex scenario.

- **Game Theory**

  Game Theory is a branch of applied mathematics, which attempts to mathematically capture behavior in strategic situations, in which an individual's success in making choices depends on the choices of others. In [40], the authors propose *bandwidth allocation algorithm* and *admission control algorithm* based on *bankruptcy game*. With this special type of N-person cooperative game, each access network cooperates to provide the requested bandwidth to a new connection using *coalition form* and *characteristic function*. The amount of allocated bandwidth to a connection in each network is obtained using *Shapley value* and the stability of the allocation is analyzed using *the core* concept. User initiating a new connection is analogous to bankrupt company and the requested bandwidth is the money that has to be distributed among different networks (creditors). The objective of each network is to offer maximum bandwidth as possible in order to gain maximum revenue from new connection, similar to creditors trying to get the most payment.

  Here is an example scenario. When a new connection requests for bandwidth, a central controller determines the amount of offered bandwidth from each network using the equation from *bandwidth allocation algorithm*:

$$d_i = \begin{cases} \tilde{b}_{k,i}, & \tilde{b}_{k,i} < (B_i^{(a)})^r \\ (B_i^{(a)})^r + \aleph\,(B_i^{(a)} - (B_i^{(a)})^r), & \tilde{b}_{k,i} \geq (B_i^{(a)})^r \end{cases}$$

where $\tilde{b}_{k,i}$ is the predefined offered bandwidth by network i to a new connection with subscription k, $(B_i^{(a)})$ is the available bandwidth in network i, $b_k^{(req)}$ is the amount of requested bandwidth in class k, $\aleph$ is a uniform random number between zero and one, and r is a control parameter which will be referred to as the bandwidth shaping parameter $(0 < r \leq 1)$. In this case, the Shapley value becomes the amount of allocated bandwidth in each network i or x_i. After that, the *admission control algorithm* ensures the requested bandwidth can be satisfied. Let C be the core, a set of stable imputations and A be the set of networks, the new connection is accepted if $\sum_{i \in A} x_i \geq b_k^{(req)}$ and $x_i \in C, \forall i \in A$ (i.e., the Shapley value is in the core, namely, the solution is stable) and it is rejected otherwise.

Discussion: Recently, game theory is gaining more popularity for solving problems in telecommunications. It has been used to model bandwidth allocation as well as pricing in the network. With the presented model [40], *coalition form* and respective *characteristic function* have to be defined appropriately. The solution is stable (i.e. everybody is satisfied) only when it belongs to *the core*, which is not always the case. In case of unstable solution, the most preferable distribution has to be determined, thus this strategy can become more expensive. We can notice here another example of joint resource management, which is a result from heterogeneous nature of the network. However, it is still unclear how to really perform integration of different network bandwidths into one connection in real scenario and this issue is not discussed in neither [39] nor [40]. Therefore, it would be interesting and beneficial to explore the feasibility of this joint connection using either simulation or experimental setup. Experimental results should be conducted in order to enhance the theoretical and numerical works.

- **Utility function**

  In economics, utility is a measure of the relative satisfaction from consumption of various goods and services; while in [41] , the authors proposed a concept of *degradation utility* to deal with different user priorities. By degrading lower priority traffic, more bandwidth can be released for higher priority users. First, network operators specify levels of service in terms of offered bandwidth (Table 2.3). Further, a classification of these services, for each application type (voice, video, and data), is marked as *excellent*, *good*, *basic*, and *rejected*. This will be used to compute released bandwidth (difference of bandwidths before and after degradation). After that, table of rewards for each user priority are defined: there

Application	**Excellent** (kbit/s)	**Good** (kbit/s)	**Basic** (kbit/s)	**Rejected**
Voice	30	30	30	0
Video	2000	384	256	0
Data	100	50	10	0

Table 2.3: Bandwidth for different quality of service.

Quality Level	**Voice**	**Video**	**Data**
Excellent	300	700	1000
Good	300	600	800
Basic	300	500	400
Forced Disconnection	-5000	-5000	-5000
Handover Drop	-5000	-5000	-5000
Reject	-2500	-2500	-2500

Table 2.4: Setting rewards for user priority class 1.

are three kinds of quality (*excellent*, *good*, and *basic*) and disconnection (*forced disconnection*, *handover drop*, and *rejected*); each of them associated with reward for each type of application (Table 2.4). This will be used to compute lost reward points (difference of reward points before and after degradation). Finally, degradation utility is the ratio of released bandwidth and lost reward points. When a new connection is requested, network operator finds all potential degradable connections, computes their degradation utilities, and begins by degrading the connection that gives the highest utility.

Here is an example scenario, consider a connection with application type: *video* and quality level: *excellent*. When the connection is degraded to *good* quality level: released bandwidth = 2000-384 kbit/s = 1616 kbit/s; lost reward points =700-600 =100; degradation utility =1616/100 =16.16.

Discussion: With tremendous growth of multimedia traffic, releasing bandwidth of low-priority traffic to give better quality for high-priority traffic becomes an interesting strategy for network operator. Degradation utility function [41] has been designed to perform this strategy but the tradeoff between satisfying upgraded connection and degraded connections has to be weighed properly. Moreover, to use this type of strategy, it is advisable to have service level agreement (SLA) signed between users and network operators in order to specify their individual responsibilities and priority class of services.

2.4.2 User-centric Approach

In this type of approach, decisions are made at user terminal and they are based only on the user's profit without considering network load balancing or other users. Therefore, the schemes in this approach mostly deal with network selection problem (including handover selection), which is to find the most profitable network for user's application. There are some debates on this approach since new users only consider their own profit and do not care about network load distribution. Thus, the network may be congested easily resulting in quality degradation of ongoing users. Furthermore, after choosing a connection, if the connection request is rejected by operator for some reasons, user will have to process again network selection resulting in higher energy consumption. In this subsection, recent techniques and their representative schemes are presented.

- **Analytical Hierarchy Process and Grey Relational Analysis**

  An Analytical Hierarchy Process (AHP) is employed for objective criteria weighting. An order preference technique based on Grey Relational Analysis (GRA) is then applied to rank the alternatives. AHP is used to solve complex decision-making problems involving different areas, including planning, resources assessment, performance measurement, resource allocation, policy selection, and priority setting. On the other hand, GRA is one of the main directions among the current applications of grey system theory, and can effectively solve the complicated interrelationships among multiple performance characteristics by optimizing grey relational grades [42].

  The authors of [43] propose to solve network selection problem using AHP to weigh QoS factors and using GRA to rank the networks. With QoS factors, the authors construct an AHP hierarchy based on their relationships similarly to Fig. 2.5. QoS is placed in the *topmost* level as the objective; main QoS factors describing network condition such as availability, throughput(α), timeliness(β), reliability(γ), security(δ), and cost (ε) are placed in the *second* level. Moreover, the authors decomposed timeliness into sub-factors delay(ζ), response time(η), and jitter(θ) and reliability into BER(λ), burst error(μ), average number of retransmission per packet(ν), and loss ratio(σ). These sub-factors are arranged in the *third* level. Finally, available solutions are arranged in the *bottommost* level. QoS parameters are separated into two types: user's preference and network conditions. User-based data is collected and processed by AHP in order to get global weights of second-level factors $GW = \{w_\alpha, w_\beta, w_\gamma, w_\delta, w_\varepsilon, w_\kappa\}$ and local weights of third-level factors $LW1 = \{w_\zeta, w_\eta, w_\theta\}$ $LW2 = \{w_\lambda, w_\mu, w_\nu, w_\sigma\}$ and then the final weights are computed $W = \{w_1, w_2, ..., w_{10}\} = \{w_\alpha, w_\beta w_\zeta, w_\beta w_\eta, w_\beta w_\theta, w_\gamma w_\lambda,$ $w_\gamma w_\mu, w_\gamma w_\nu, w_\gamma w_\varepsilon, w_\delta, w_\varepsilon\}$. At the same time, network-based data are normalized by GRA, and the ideal network performance is defined following by calculation of the grey relational coefficient (GRC) which gives grey relationship between ideal network and the other. The calculation of GRC is expressed as

$$GRC_{UMTS/WLAN} = \frac{1}{\sum_{P=1}^{10} w_p |x^*_{UMTS/WLAN}(p) - 1| + 1}$$

where $x^*_{UMTS/WLAN}(p)$ is the normalization of the UMTS data or the WLAN data. The network with the largest GRC is the most desirable.

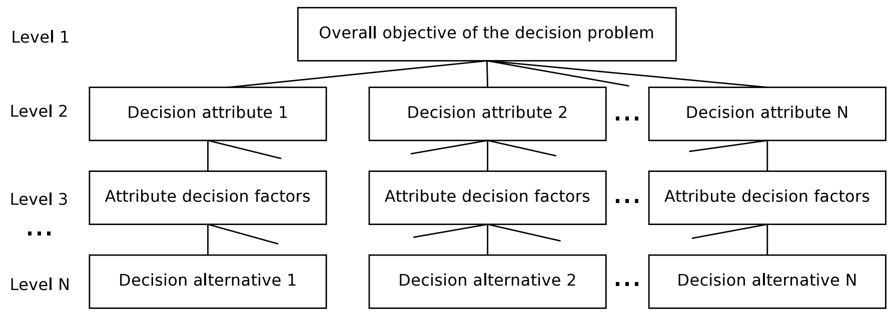


Figure 2.5: Structuring analytical hierarchy process [44].

Discussion: After this mechanism [43], AHP and GRA are also deployed in other network selection mechanisms [45, 46] or scheduling [47] as well. We can notice that Multi-Attribute Decision Making (MADM) [48] has recently gained popularity in telecommunications as it is suitable to complex decision making problem today. In fact, MADM refers to making preference decisions (e.g. evaluation, prioritization, selection) over the available alternatives that are characterized by multiple, usually conflicting, attributes. Other than AHP and GRA, which are among the most popular MADM algorithms, TOPSIS (Technique for Order Preference by Similarity to Ideal Solution), MEWS (Multiplicative Exponent Weighting), and SAW (Simple Additive Weighting) are also deployed in decision making under heterogeneous environment.

- **Consumer surplus**

  In economics, the consumer surplus is the amount that consumers benefit by being able to purchase a product for a price that is less than they would be willing to pay. In [49], the authors propose a user-centric solution using *customer surplus* concept for network selection in HWNs. The scheme has been designed for non real-time traffic with the following strategy. First, the users survey the radio interface and determine a list of available access networks. Next, they predict the transfer completion time (T_C) of each available access network on the list according to $T_C = F_i/r$; where F_i is size of file i in bits and r is average rate for total transfer in bps. The average of the last five data transfers is used to derive the completion time. After that, they compute predicted utility $U_i(T_C)$, which is the relationship between the budget and the user's flexibility in the transfer completion time. For each candidate network, the user computes consumer surplus (i.e., $CS = U_i(T_C) - C_i$ subject to $T_C \leq T_{Cmax}$, where T_{Cmax} denotes the maximum transfer completion time that a user is willing to wait). In other words, CS is the difference between utility and cost (C) charged by the network. Finally, the best network (giving maximum CS) is chosen.

Discussion: This scheme has been designed for non real-time application and it is not appropriate for today's real-time multimedia application, which relies on more than only completion time. However, the concept of customer surplus remains interesting as it can also be exploited using different parameters in real-time multimedia applications. For example, we can collect reliable information on the quality of access networks with technical support from IEEE 802.21 standard [31], and can possibly combine this information in a dynamic decision mechanism in order to deduce the cost of each network (incentive for user selection).

- **Profit function**

  In economics, a profit function is defined as $\pi(p,w) = max_x p f(x) - wx$, where w and x are vector of factor prices and factor demands respectively, p is the output price [50]. The profit function maps particular factor prices to the maximum profit levels achievable at those output prices and factor prices. In [51], the authors took a slightly different definition to handle handoff selection in HWN. They associate each handoff with a *profit* that is decided by a target function with two parameters: bandwidth gain and handoff cost. Moreover, they classified handoffs into *reactive* and *proactive* handoff. A reactive handoff is initiated whenever a mobile node is going to roam out of the current cell, while proactive handoff can only be initiated at periodical discrete epoch when connection experience can be improved.

  Their profit function is defined as $P = f(G,C)$. The *bandwidth gain G* gives the difference in bandwidth between the next period and this period. Its definition of a handoff decision at epoch t_k is defined as

$$G_i(t_k) = \begin{cases} m(i,t_k) - m(j,t_{k-1}) & k \geq 1 \text{ (handoff connection)} \\ m(i,t_k) & k = 0 \text{ (first-time connection)} \end{cases}$$

  where i, j are network indexes, $i \neq j$ means an inter-system handoff (proactive or reactive), $i = j$ either means an intra-system handoff (proactive) or no handoff; $m(i,t_k)$ is the bandwidth of network i used by mobile node between two handoff decision epochs $[t_k,t_{k+1})$. The authors define the *handoff cost* as data volume lost due to handoff delay; it corresponds to the volume of data which could have been transmitted during the handoff delay. Its expression is $C(t_k) = m(i,t_{k-1})d(x,y)$

where $d(x,y)$ is the handoff delay when a mobile node makes a handoff from base station x to y. Thus, the profit is a difference between gain and cost. At each handoff epoch, mobile node compares profit from different networks and chooses the one that yields maximum profit

$$P_i = (t_{k+1} - t_k)G_i(t_k) - m(i, t_{k-1})d(x,y).$$

Discussion: Similar to previous concept of customer surplus, this profit function [51] compares gain and cost to obtain utility of candidate networks. The reactive handoff is designed for moving user that needs to select rapidly the best network from its neighborhood whereas the proactive is for improving quality of service when a better network is present in the neighborhood. This adaptation is interesting in wireless network with changing condition and user mobility.

2.4.3 Collaborative Approach

Besides two previously described approaches, a collaborative approach is the most compromising in terms of profit between users and network operator since it takes into account the profit of both sides for making decisions. Moreover, since both network operator and user participate in resource allocation, the problem of connection rejection as in user-centric approach will not occur. Recent techniques and their representative solutions are presented in this subsection.

- **Fuzzy Logic Controller**

  The authors of [52] use an algorithm based on *fuzzy logic controller (FLC)* to evaluate fitness ranking of candidate networks. At first, they differentiate decision making into three phases: *pre-selection*, *discovery*, and *decision making*. Pre-selection phase takes criteria from user, application, and network to eliminate unsuitable access networks from further selection. If available networks are not corresponding to user's requirement, system returns to ask the user to reduce their criteria. The discovery stage deals with two kinds of state: *power-up* users (when no current connections exist), and *connected* users (when a connection is already established but QoS is not meeting the criteria at the same time other potential networks become available). The authors implemented discovery phase based on fuzzy logic control, they fuzzify crisp values of the variables (network data rate, Signal to Noise Ratio, and application requirement data rate) into grade of membership in fuzzy set. Then, these membership functions are used as input to the pre-defined logic rule base. Finally, overall ranking is obtained through defuzzification with weighted average method.

  Discussion: After its first application in handoff management by the authors of [53], fuzzy logic control is becoming popular again in HWN management as many schemes (e.g. [52, 54, 55]) have been proposed recently. The current scheme [52] deployed it for network selection; FLC gives a good result in this case of few metrics. However, if the number of metrics increases, the system may become very complex and may give erroneous results. The critical issue in this approach is the definition of fuzzy set and rules which needs to be carefully specified. These specifications are very important in order to get a good approximation and they are very delicate to define.

- **Objective function**

  Objective function, or goal function, is the function to be optimized, depending on the object parameters. It constitutes the implementation of the problem to be solved. The authors of [56] applied this concept to network selection in HWN. In their objective function, inputs are derived from three different sources: *user data*, *network data*, and *policy information*. First, users are asked for a list of visible access points (AP) with corresponding signal quality, a list of requested services with corresponding nominal bit rate, and delay tolerance. Second, network data, such as the AP bandwidth and the delay of the queue between access router and the backbone, are collected. Third, policy such as cost, compatibility, trust, preference, and capability along with their weights are defined. The weights can be dynamically changed according to the network condition. Finally, with all factors and their weights, the algorithm iterates and computes the best allocation that maximizes the objective function for overall network.

  For the access and interface selection algorithm, the authors denote requested service as s belonging to the total of services S and ap is access point belonging to the total of access points AP, the objective function is then

  $$OF(\forall s \in S, \forall ap \in AP) = F(s, ap) + OF(\forall s' \in S, s' \neq s, \forall ap \in AP).$$

  The value of the OF for s' represents the allocation of the rest of services. The sequence by which the OF is calculated affects the overall result, because the allocation of an application to an AP decreases its available bandwidth. Thus, all possible permutations must be considered. Function F consists of the quality part Q and the part of policies PT, with their corresponding weights $(w_q + w_{pt} = 1), F = w_q Q + w_{pt} PT$. While Q and PT are analyzed as $Q = w_{bi} BI + w_{di} DI + w_{sqi} SQI$ and $PT = w_{cci} CCI + w_{npi} NPI + w_{tti} TTI$. Note that $w_{bi} + w_{di} + w_{sqi} = 100$ and $w_{cci} + w_{npi} + w_{tti} = 100$. BI is bandwidth indicator, DI is delay indicator, SQI is signal quality indicator, CCI is cost and compatibility indicator, NPI is network provider indicator, and TTI is terminal type indicator.

  Discussion: This scheme includes all necessary factors to make a good decision. Moreover, it also proposes to use multihoming for implementation of joint bandwidth allocation. The main actor who makes decisions in this scheme is the network operator; however, the scheme also requests for lots of information from the user raising transparency and feasibility issues in real implementation.

- **TOPSIS**

  The principle behind TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) is described in [57]: the chosen alternative should be as close to the ideal solution as possible and as far from the negative-ideal solution as possible. The ideal solution is formed as a composite of the best performance values exhibited (in the decision matrix) by any alternative for each attribute. The negative-ideal solution is the composite of the worst performance values. Proximity to each of these performance poles is measured in the Euclidean sense (e.g., square root of the sum of the squared distances along each axis in the attribute space), with optional weighting of each attribute. The authors of [58] proposed an algorithm for path selection on multihomed end-hosts based on TOPSIS. In this mechanism, the authors collected parameters from both network level (QoS parameters: bandwidth, delay, jitter, and BER) and application level (traffic class: conversational, streaming, interactive, and background).

  The authors deployed TOPSIS for their Score Calculator. The first step consists in formatting the data in a matrix X_{ij}, of which each row represents the measurement of parameters of a path. The authors normalized the value of every parameter using $n_{ij} = x_{ij}/\sqrt{\sum_{j=1}^{m} x_{ij}^2}$. Then, each column of the matrix is multiplied by the corresponding weight w_i using the formula $v_{ij} = w_i * n_{ij}$ and $\sum_{j=1}^{n} w_i = 1$. These weights are deduced from the QoS class of the relevant application in [59]. Next, the authors extract the ideal points (negative and positive) from the normalized weighted matrix V_{ij}

  $$A^- = \{v_1^-, ..., v_n^-\} = \{(min(v_{ij})|i \in I), (max(v_{ij})|i \in I)\}$$

  $$A^+ = \{v_1^+, ..., v_n^+\} = \{(max(v_{ij})|i \in I), (min(v_{ij})|i \in I)\}$$

  After that, distances of each alternative to the two ideal alternatives are computed as $d_j^+ = \{\sum_{j=1}^{m}(v_{ij} - v_i^+)^2\}^{1/2}$ and $d_j^- = \{\sum_{j=1}^{m}(v_{ij} - v_i^-)^2\}^{1/2}$. Finally, the score of each alternative j is computed as $R_j = d_j^- / (d_j^+ + d_j^-)$. These scores are used in the flow distribution process.

  Discussion: Similarly to AHP, TOPSIS is one of the MADM techniques. It has been deployed here for path selection of multi-homed node. TOPSIS is easy to use as its software is available for the implementation. However, according to vertical handoff comparison [59], the performance of TOPSIS is slightly lower in bandwidth and in delay than GRA for interactive and background traffic.

For a better comprehension, the surveyed schemes are presented in Table 2.5. They are arranged in terms of deploying technique, input parameter, procedure, output, approach, and joint allocation (whether the scheme assumes joint resources or not).

Techniques	Parameters	Procedure	Output	Approach	Joint allocation
SLP	Allocation, demand, underutilization, and rejection	1-association of predetermined probability to demands 2-variable formulation 3-SLP statement	Allocation in each network	Network-centric	Yes
Game Theory	Available bandwidths in each network	1-determine offered bandwidths 2-compute Shapley value 3-verify core	Bandwidth allocation	Network-centric	Yes
Degradation Utility	Released bandwidth and lost reward point	1-compute ratio of released - bandwidth & loss reward point for each connection 2-find maximum	Connection that gives maximum utility	Network-centric	No
AHP & GRA	User's requirements and network conditions	1-AHP of user's requirements 2-GRA of network conditions 3-compute GRC	Network rank by GRC	User-centric	No
Consumer Surplus	Utility and cost	1-compute the difference between utility and cost for each network 2-find maximum	Network that gives maximum benefit	User-centric	No
Profit function	Bandwidth gain and handoff cost	1-compute the difference between gain and cost for each network 2-find maximum	Most appropriate network for handoff	User-centric	No
FLC	Network data rate, SNR, application - required data rate	1-fuzzification 2-fuzzy inference 3-defuzzification	Fitness rank of each network	Collaborative	No
Objective function	Quality and policy indicators	1-compute sum of (inputs$\times$weights) for each network 2-find maximum	Allocation of services to APs and terminals	Collaborative	Yes
TOPSIS	QoS parameters and traffic class	1-format data into normalized matrix 2-compute data$\times$their weights 3-compute ideal points (+/-) and distances from ideal points 4-select the best solution	Best path for flow distribution	Collaborative	No

Table 2.5: Summary of the surveyed schemes.

2.5 Related Issues

Although decision mechanism is essential in the RRM framework, other supports in terms of QoS and mobility are necessary to handle the variety of applications in mobile terminals today. Moreover, in order to have an efficient scheme, architectural design should also be considered for system performance and end-server or end-user can participate in resource management using media adaptation.

2.5.1 QoS Support

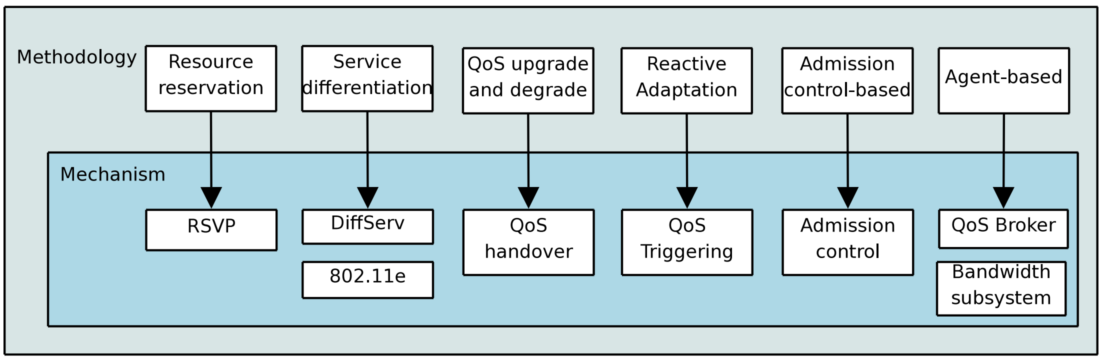


Figure 2.6: QoS Support: Methodology and Mechanism.

One of the most important issues to consider when designing RRM framework is QoS. Practically all network operators aim to guarantee the best connections to users. In this context, there are many RRM schemes that take QoS metrics as well as user requirements into account for decision making. Some schemes ([60] and [61]) make use of *resource reservation* protocol to pre-reserve resource and to guarantee requested quality. Many others have been proposed but most of them fail for deployment since all network equipments have to support reservation. In addition to resource reservation, *service differentiation* has also been used to distinguish treatments for applications with different priorities. Differentiation of service can be done at different levels (IEEE 802.11e at MAC layer, [52] and DiffServ (Differentiation of Service) at the network level) by means of priority-schedulers that help dealing with requests according to their priorities.

As in wireless environment, users are mobile and can move from one place to another while being connected. The authors of [51] apply this movement to improve QoS. They deploy *QoS handover*, a type of handover aimed to improve quality. However, QoS handover induces delay which results in packet loss. Related technique is *QoS upgrade/degrade* proposed in [41]. Utilization of this mechanism has to be carefully studied beforehand due to tradeoff between degrading and upgrading connections. When QoS upgrade takes place, someone is being degraded to release necessary bandwidth (in case of saturated network). Nevertheless, this approach is interesting because it provides suitable solution for increasing problems of multimedia traffic; for example, network operator can upgrade delay-sensitive traffic (multimedia class) while degrading delay-insensitive traffic (background class).

In addition to the techniques previously described, there are new architectures that have been designed for supporting QoS. Most of them use agents called *QoS broker* [60, 62] in order to manage QoS in the network. Controlling QoS can be done periodically as in [51], but high control signaling wastes bandwidth; particularly in the case of limited-bandwidth network such as GSM or GPRS, where control traffic introduces bottleneck point in the network. To avoid this problem, dynamic adaptation using *triggering* seems to be more adequate. Triggering conditions depend on network operator's objective; for example, according to [56], system is triggered when new connection arrives or when ongoing connection faces QoS problem. To cover all aspects of QoS, a framework has been proposed in [63] with three planes management providing both static and dynamic QoS functions. Finally, admission control mechanisms can also be used to support QoS by filtering new connection to maintain QoS level of ongoing connections.

2.5.2 Mobility Support

As mentioned earlier, stations in HWNs are mobile and can move freely from one place to another. To handle this mobility, many works have been proposed using mobility management modules. Most of them are managed at network providers using *Mobile IP (MIP)* [64] or its extension such as Fast handover for MIP or Hierarchical MIP. Moreover, some works ([65] and [66]) proposed mobility support using Session Initiation Protocol (SIP) at application layer.

Other works have focused more on the *handover process* itself with the objective of achieving seamless handover. Authors in [67] discussed on the detail of handover by proposing a function that determines the best handover initiation time in order to avoid early or late initiations. Early initiation will result in double use of bandwidth in home and foreign networks while late initiation will result in packet loss and non-seamless handover.

A technique like *multi-homing* has also been used to improve performance in mobile networking as in [56] and [58]. With multi-homing, it is possible to connect to multiple networks at the same time using multi-interface terminal. Advantages of such solution are the decrease in handover delay and more reliable connection in case of link failure but the drawback is the multiple bandwidths occupied by multi-homed terminal.

As for support on mobility management architecture, the authors of [68] have proposed a middleware named *Ubique architecture*. It allows mobile terminals to automatically select the best interface for each application flow while taking into account various requirements. More interestingly, the authors of [69] studied mobility support in an interesting way; they propose to give network operator a possibility to implicitly influence decision made by multi-interface user. For that, operator can play with assigning different weights to a set of parameters (Bit-Rate, Error, Delay, Cost, and Security).

To standardize handover, IEEE working group is developing *IEEE 802.21* [31] for media independent handover services, which will enable co-operative handover decision making of users and operators. It can be noticed that huge effort has been put on mobility issue because this issue will obviously result in quality of a service, the goal of both provider and user.

2.5.3 Architectural Design

Management architecture can be classified into three types according to how network entities communicate among each other. A *centralized* architecture is the first architecture to be discussed. Control in this architecture is aggregated into one central point usually situated in the core network as illustrated in Fig. 2.7a. Examples can be found in [39, 61, 52, 43]. Central node has a global view of the whole system, which allows an advantageous management of overall performance. However, since management is centralized at one point, all other nodes have to send management traffic to this point and this may waste bandwidth and cause congestion in the access network with limited-bandwidth capacity. Moreover, centralized architecture is not scalable and results in one point-of-failure problem.

Unlike the centralized architecture, control in *distributed* or *decentralized* architecture is delegated to several entities as illustrated in Fig. 2.7b. In general, the control is placed at access router [56] when network provider wants to manage the whole access network. Alternately, control may also be placed at the point of attachment that represents local cell such as access points or base stations. Occasionally, distributed architecture placed control on user's terminals in order to get information from user. Some solutions ([49] and [51]) give user possibility to make decision on which network to be connected. This approach is not recommended because it may result in load balancing problem since users only consider their benefits without considering actual load in the target network. Moreover, when the decision is made by user, it does not imply that the connection will be successfully accepted by the selected network operator, who may prefer to reject less-valuable call to accept another more-valuable one. In addition to these distributed approach, the authors of [41] proposed cooperative distributed system to manage the whole heterogeneous system while still being scalable.

The last approach is a *hybrid* architecture, which combines the two architectures described above. It is composed of a central node that manages global resource and distributed nodes to manage resource locally (Fig. 2.7c). We also observe schemes collaborating management in distributed network node as well as user terminal. For

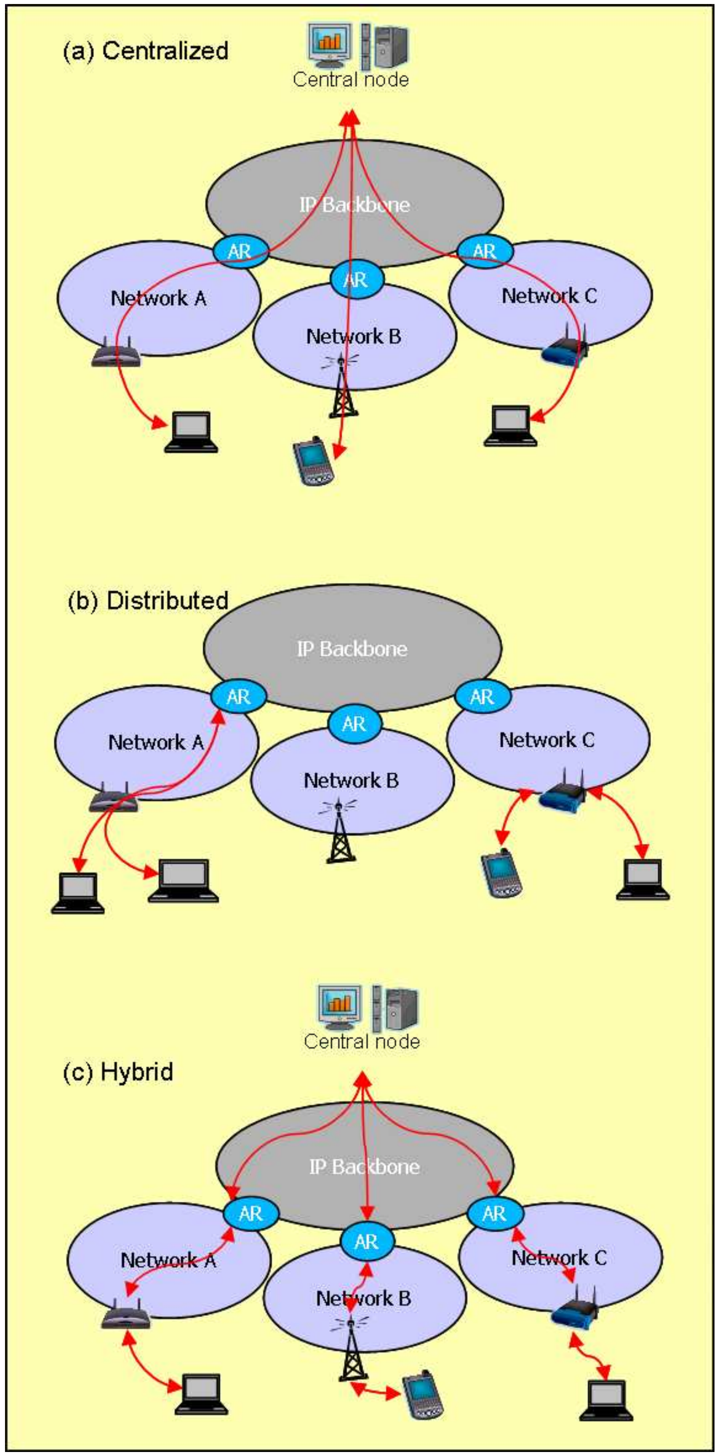


Figure 2.7: Different Types of Architecture.

example, the authors of [70] recommended the combination of distributed network and terminal management for dynamic handling of individual users and sessions. In [56], the authors presented network-based and terminal-assisted approach to optimize resource allocation while compromising QoS constraints. The authors of [67] developed a hybrid network selection scheme that combines terminal-based and network-based selection mechanisms. In this scheme, terminal dynamically collects network condition and determines best reachable network, then network makes globally optimized selection and achieves load balancing for the whole system.

2.5.4 Media Adaptation

In today's wireless environment, multimedia traffic such as video transmission increases considerably. With this kind of traffic and unstable condition of wireless network, media adaptation becomes essential. Media adaptation means that node adapts itself to media condition. For example, the control of encoding rate of the video stream based on the estimated available bandwidth or the error correction according to the varying wireless conditions.

Media adaptation can be performed at different locations: end systems or intermediate nodes. End systems such as sender or receiver may participate in media adaptation. The sender can adapt its parameters to be coherent with network condition and ongoing application. For example, the server adjusts its transmission rate according to congestion in the network. *Stream switching* is one of the techniques. The server prepares streams to be transmitted to the channel in different encoding rate and stocks them in a database. When network condition changes; the server selects stream with encoding rate accordingly. However, drawback of this technique is high consumption of disc space that cannot be possible in every case. It can be noticed that sender adaptation is optimized in terms of signaling since no bandwidth is used for communication between sender and receiver. Receiver can also cooperate in dynamic adaptation by sending its reception capacity to sender but this approach may be costly in terms of communication. So it is not recommended in small-bandwidth networks such as GPRS. More recently, *scalable video coding (SVC)* [71] has been released. With this technique, encoding rate can change dynamically according to network condition using concept of base and enhancement layers. With SVC capabilities, the authors of [72] have proposed a context-aware video delivery and an architecture [73] for service and mobility management.

Another issue in media adaptation is *reliability*. To deal with unreliable channel, error correction mechanisms are also recommended. For example, forward error correction (FEC) and automatic repeat request (ARQ) have been deployed in [74] to enforce transmission. However, for real-time or delay sensitive application, ARQ is not appropriate because late arrival of retransmitted packets are usually discarded. To deal with retransmission, the author of [75] has proposed selective retransmission scheme to adaptively enable retransmission according to channel condition. The retransmission should be disable when the channel is congested otherwise it should be enable with selective retransmission of I, P and B frames[1]. Finally, the authors of [60] proposed an complete architecture with adaptation in different levels; for example, channel adaptation module using several protocol such as H264/AVC to provide enhanced bit error resilience capability, UDP-Lite (RFC3828) to deliver erroneous packets and to deal with erroneous packet payloads, robust header compression (RoHC) to reduce IP overhead improving IP packet latency for real time services, and finally FEC to eliminate retransmission that degrade overall throughput.

2.6 Conclusions

Research in radio resource management has been extensively studied in recent years, and many schemes have been proposed. In this chapter, solutions for decision mechanism are investigated in details and they are classified into three approaches: network-centric, user-centric, and collaborative. There are a few trends in decision making under heterogeneous environment. Popular techniques are (i) *economics-like* functions that compute the benefit and cost in order to derive the best solution such as profit function, degradation utility, or customer surplus; (ii) *mathematical methods* such as game theory, stochastic programming, and objective function; and (iii) *multi-attribute decision making* such as Fuzzy Logic Controller (FLC), Analytical Hierarchy Process (AHP), or Technique for Order Preference by Similarity to Ideal Solution (TOPSIS).

A new issue raised in heterogeneous environment is *Joint Resource Management*, in which bandwidth allocation to a user can be provided by different networks simultaneously. This idea is interesting in HWNs because allocated bandwidth to a user can be provided by several networks and thus problem of load-balancing can be alleviated. However, it is still questionable how to set up this type of integrated connection in real scenario. Besides, it would be complicated to handle billing or authentication issues, not only at users but also among network operators themselves; a fine-grained study has to be conducted before this type of solution can be released to the market.

In order to design an efficient mechanism, this chapter also discusses QoS and mobility supports that arise due to the emergence of multimedia in wireless environment. These two issues are influencing the research and development in wide areas, and they need to be considered when a new scheme is designed. Moreover, there has been an ongoing debate on architectural design in terms of system performance; and finally, a hybrid scheme is recommended for good performance of the system because network operator can have a global view of the system while still being scalable. The latest trend in multimedia network management also includes media adaptation, in which

[1] In MPEG encoding, three frame types are used to represent the video: Intra (I), Predicted (P), and Bi-directional (B).

the end point like multimedia server can adapt their capability according to current network condition or with collaboration from end-user.

This chapter has described state of the art in resource management based on quality. It can be noticed that heterogeneity in network technology has brought great advantages into the service. However, a main difficulty is how to collect information from different technologies along with their various parameters in order to make management decision. In the next chapter, reader will be introduced to another concept of quality called *Quality of Experience*. Its definition and comparison to the well known quality of service or QoS will be explained. Methods for its assessment will also be investigated and the question of how this concept can be used in real-time resource management will be discussed. The main interest of studying quality of experience is the fact that it is independent of network technologies and applications and hence can be used as a context-independent metric for managing heterogeneous networks today.

Chapter 3

Quality of Experience in Resource Management

3.1 Introduction

As multimedia applications have emerged, representing quality of a provided service using technical or QoS parameters is no longer suitable. Multimedia traffic should not be measured in terms of throughput, loss rate, or delay but more in terms of user experience such as good or bad. Therefore, this chapter presents a new concept of service quality, which is expressed in terms of user satisfaction or *quality of experience*. The focus is on wireless network environment, which is gaining tremendous success nowadays. The need of quality-based resource management in this type of environment is crucial, especially with multimedia applications. Network operators wish to control their resources efficiently while maintaining user satisfaction. At the same time, traditional ways of managing network, using information from monitoring technical parameters, fail to give accurate evaluations of user experience; hence the inspiration of the study in this chapter.

The rest of this chapter is organized as follow. At the beginning, Section 3.2 gives definition of quality of experience and its comparison to the well-known quality of service. Then presentation of different assessment approaches and their performance evaluation is given in Section3.3. A comparison of assessment approaches is presented to provide a better comprehension of QoE measurement. The focus will be on the hybrid technique called PSQA (Pseudo-Subjective Quality Assessment) that keeps advantages and avoids drawbacks of the other approaches. It enables accurate and real-time resource monitoring and management. Management possibility with QoE is discussed in Section 3.4. Finally, section 3.5 gives conclusions.

3.2 QoE in Network Evolutions

With network evolution nowadays, quality becomes a critical factor as it drives economics in many ways, e.g. service level agreement, quality differentiated services, and billing. The bottom-line of quality is customer satisfaction, which is a combination of many factors such as availability, quality, price, or utility. As a consequence, resource management must be done in real time and must address user experience or currently called quality of experience. In this section, backgrounds on network evolution, definitions, and comparison between QoE and QoS is provided.

- **Network Evolution**
  Before multimedia communication era, QoS parameters were enough to evaluate quality of provided services. However, as today's real-time multimedia applications emerged and users became more experienced; emphasis has then shifted from packet level to service or user level. Existing metrics are no longer enough because simple network statistics will not reflect user's perception. For example, loss rate, a widely used quality indicator, is not always reliable when dealing with quality of experience. In fact, high loss does not automatically imply bad perception. If sender uses preventive technique like FEC, QoE can be maintained at acceptable level. Moreover, the content of the media also plays an important role as the same loss in a soap opera may not have the same perceptual effect as in a football match.

  As for network technology, wireless network is taking place of wired network progressively giving birth to wireless multimedia network or WMN. This phenomenon has pushed the number of Internet users through significant increase. Besides, wireless nature (i.e. limited bandwidth, shared resources, channel interference...) is easy to be over-utilized. Network load must be controlled carefully so that acceptable quality is maintained while network operators are not penalized with underutilization. In general, to guarantee good perception at users, an IP triple-play broadband operator should always ensure that every link in backbone will transport less than the 50% of its capacity. This is to prevent congestion in case of failure of a redundant link.

  The purpose of this dissertation is to avoid such a conservative approach by studying possibilities and performances of using QoE as metric for managing resources. This new paradigm will provide a better flexibility while maximizing throughputs and keeping consistent perception at users. Later, the demonstration will focus on two representative wireless technologies: WLAN (the widespreading) and UMTS (supporting high mobility). In addition to homogeneous environment, a study is also conducted for heterogeneous system composing of these two technologies. Similar ideas can also be applied in the future to other network technologies or other heterogeneous network environment.

- **QoE Definition**
According to ITU [9], *Quality of Experience* is the overall acceptability of an application or service, as perceived subjectively by the end-user. In other words, QoE is a subjective concept representing the actual quality of a service perceived by end-user. It can be rated in terms of user impression of the service such as good, fair, or bad. As a measure of user satisfaction, QoE is an important metric for the design of systems. As such, it is also an indicator of how well the system meets its targets. This is particularly relevant for multimedia services because bad network performance may affect drastically user experience. Therefore, expected QoE is often considered as a system output metric during the design. This QoE metric is often measured at the end device; however, the overall acceptability may be influenced by user expectations and context.

- **QoE versus QoS**
QoE differs from the well-known QoS in many ways. QoE is a measure of *end-to-end* performance at the service level from the user perspective while QoS is a measure of performance at the packet level from the network perspective. More precisely, QoE is *subjective* and relates to the actual perceived quality of a service by the user, this applies to voice, multimedia, and data, whereas QoS is an optimization tool designed to deliver a certain quality of experience by ensuring that network elements apply consistent treatment to traffic flows as they traverse the network. QoE can be used to describe the performance of a device, system, service, application (or their combination) from *user point of view*. For networking aspect, QoE measures how well a network service satisfies users' expectations and needs. On the other hand, QoS refers to a set of technologies (QoS mechanisms) that enables the network administrator to manage application performance. In other words, QoS mechanisms help to manage available bandwidth more efficiently. Finally, QoE is *context-independent*; quality expressed with QoE (e.g. "good") has the same meaning in all technologies and applications. On the contrary, QoS is context-specific different technologies and applications may have different QoS parameters. Table 3.1 summarizes differentiation between QoE and QoS in terms of performance, OSI layer, perspective, concept, and context-dependency.

In terms of	**QoE**	**QoS**
Performance	End-to-End	Packet level
OSI layer	Session and upper	Network and lower
Concept	Subjective	Objective
Perspective	User	Network
Context-dependency	Independent	Dependent

Table 3.1: QoE vs. QoS.

3.3 QoE Assessment

Before being capable to use QoE in network management, an appropriate QoE assessment tool is needed. As mentioned earlier, QoE is a subjective concept; hence, QoE assessment is a difficult task. QoE does not depend entirely on video and network quality but it also depends greatly on user's opinion and experience. In addition, test environment (including screen size, monitor resolution, luminance, etc.) also plays an important role. As an example, quality of a video on YouTube page seems to be excellent when watching from small embedded window on the page but it is much worst when enlarging to full-screen mode. Therefore, many techniques have been developed in order to assess as accurately as possible this perceptual quality. To investigate QoE measurement, this section presents three approaches namely *subjective approach*, *objective approach*, and *hybrid approach*. It also presents performance evaluation of these approaches for assessing QoE in video streaming application over wireless networks under different conditions. More specifically, a hybrid approach called *PSQA* is the focus because it keeps advantages of both subjective and objective schemes whilst minimizing their drawbacks.

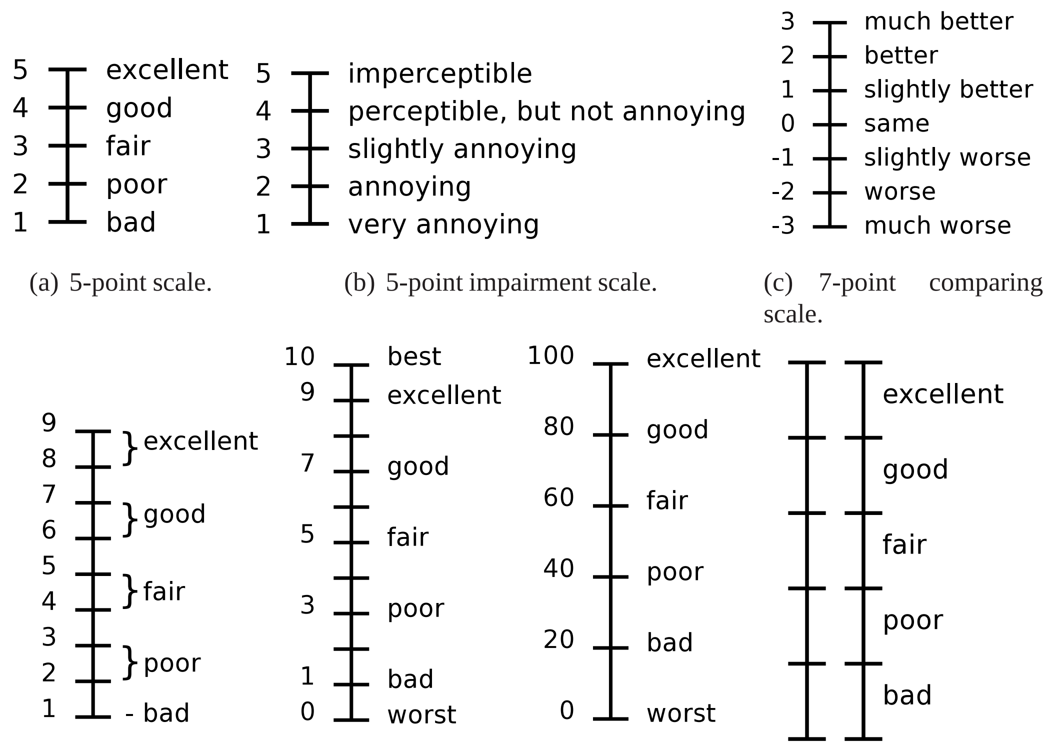
3.3.1 Different Assessment Approaches

For a better understanding of QoE, this section will give an overview of different approaches used to measure QoE, ranging from traditional subjective approach through objective and hybrid approach. Furthermore, their performances are compared in order to select the most appropriate method for the study.

- **Subjective Approach**

  The most accurate approach to assess perceived quality is the subjective assessment because there is no indicator of quality more accurate than the one given by human user. However, the quality score given by a human also depends on his/her own experience. The assessment consists in building a panel of human observers, which will evaluate sequences of video depending on their point of view and their perception. Quality of experience can be expressed in terms of user satisfaction, as presented in Fig.3.1.

  Standard methods for conducting subjective video quality evaluations are given in ITU-R BT.500-11 [10]. Its variations are Single Stimulus (SS), Double Stimulus Impairment Scale (DSIS), Double Stimulus Continuous Quality Scale (DSCQS), Single Stimulus Continuous Quality Evaluation (SSCQE), Simultaneous Double Stimulus for Continuous Evaluation (SDSCE), and Stimulus Comparison Adjectival Categorical Judgment (SCACJ). All the variations are pretty much similar; changes concern for example, evaluation scale, video reference, video sequence length, evaluation scale, number of video per trial, or number of observer. To



(a) 5-point scale. (b) 5-point impairment scale. (c) 7-point comparing scale.

(d) 9-point scale. (e) 11-point scale. (f) continuous scale. (g) double continuous scale.

Figure 3.1: ITU standard scales for subjective test methods [76].

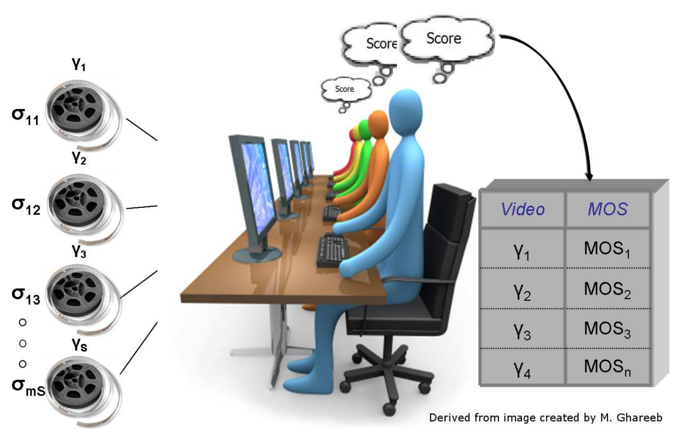


Figure 3.2: Subjective Evaluation Campaign.

conduct appropriate subjective assessments, it is necessary to select from different available options those that best suit the objectives and circumstances of the assessment problems.

Despite that subjective approach is the most accurate; it is very expensive in terms of time and manpower. Moreover, the assessment process is very complex and has strict requirements. Therefore, it cannot be used in an automatic measurement or real-time monitoring tool. For this chapter, *Single Stimulus or SS* method will be tested; in SS, a single sequence of video is presented one at a time and the assessor provides a score for each presentation (as shown in Fig.3.2). The final score for each video sequence is the average of all observers, excluding bad observers (filtered out by a statistic filter).

- **Objective Approach**

  Since the subjective approach is not appropriate for implementation, many researchers have been looking for another approach that can be processed automatically using information such as network parameters. Consequently, they came up with an objective approach that uses algorithms or formulas and quality of service measurements of a stream given by technical parameters that can be collected easily from the network. Many objective metrics exist such as Peak Signal to Noise Ration (PSNR), ITS' Video Quality Metric (VQM), EPFL's Moving Picture Quality Metric (MPQM), Color Moving Picture Quality Metric (CMPQM), Normalization Video Fidelity Metric (NVFM). For the study, PSNR [11] is selected because it is the most common and simple objective video quality assessment widely used by many researchers. PSNR is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. It is defined via the Mean Squared Error or MSE between an original frame o and the distorted frame d as following:

$$MSE = \frac{1}{M.N}\sum_{m=1}^{M}\sum_{n=1}^{N}|o(m,n)-d(m,n)|^2$$

  where each frame has $M \times N$ pixels, and $o(m,n)$ and $d(m,n)$ are the luminance pixels in position (m,n) in the frame. Then, PSNR is the logarithmic ratio between the maximum value of a signal and the background noise (MSE). If the maximum luminance value in the frame is L (when the pixels are represented using 8 bits per sample, $L = 255$) then:

$$PSNR = 10\log\frac{255^2}{MSE}$$

  It can be noticed that PSNR can be computed only once the image is reconstructed at the receiver, hence, it may not be appropriate to use in real-time mechanisms. This is one disadvantage of such metric. The other would be the reliability to derive user experience from this metric. However, according to [12] there exist heuristic mappings of PSNR to Mean Opinion Score (MOS) as shown in Table 3.2.

PSNR [dB]	MOS
> 37	5 (Excellent)
31-37	4 (Good)
25-31	3 (Fair)
20-25	2 (Poor)
< 20	1 (Bad)

Table 3.2: Possible PSNR to MOS conversion.

- **Hybrid Approach**

  Apart from two precedent approaches, a hybrid approach tries to provide a compromise solution between subjective and objective approach. It can be noticed that many standard methods have been proposed to assess quality of experience for VoIP application, for example, E-model from ITU G.107 [13], Perceptual Speech Quality Measure (PSQM) and measuring normalizing block (MNB) from ITU P.861 [14], or Perceptual Evaluation of Speech Quality (PESQ) from ITU P.862 [15]); however, very few exists for video streaming applications. Examples are V-Factor [77], k-dimensional Euclidean space approach, and Pseudo-Subjective Quality Assessment [78].

  For the study, the *PSQA* technique is selected because it provides accurate QoE assessment with ease of use. The PSQA method is based on statistic learning using random neural network. It is hybrid in the sense that there is somehow a subjective evaluation in the methodology but this can be done only once and used as many times as necessary with the help of quality factors (objective parameters). Further descriptions will be explained in the following section.

3.3.2 Pseudo-Subjective Quality Assessment

Pseudo subjective quality assessment or PSQA is based on statistic learning using random neural network as briefly explained in the following.

- **Random Neural Network**
  Random neural network (RNN) is a simplified model of a biological nervous system. The RNN is formed by a set of neurons which exchange signals in the form of spikes, like the natural pulsed behavior. Each neuron's state is a non-negative integer called *potential*, which increases by 1 when a positive signal (an

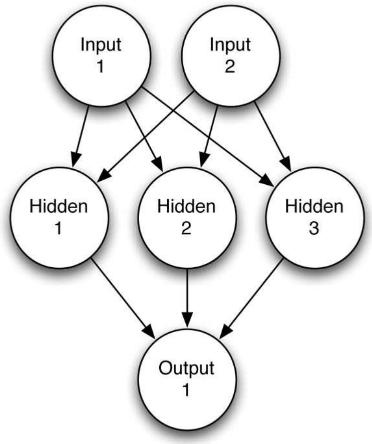


Figure 3.3: Typical feedforward neural network with a single hidden layer [79]. Similar to other types of neural networks, it begins with an input layer, which may be connected to a hidden layer or directly to the output layer. There can be any number of hidden layers, as long as there is at least one hidden layer or output layer provided.

excitation) arrives to it, and decreases by 1 when a negative signal (an inhibition) arrives. The signals can originate outside the network, or they can come from other neurons. Usually, the signals cannot flow in an arbitrary way: a topology is designed that specifies which neurons receive signals from which ones, and which neurons receive signals from the environment.

In the PSQA methodology, a particular RNN architecture is used, where each neuron behaves as a $./M/1$ queue with respect to positive signals. This means that positive signals are interpreted as customers; these customers arrive to the neurons, and are served in a FIFO (First In First Out) order. The service rate at neuron i is denoted by μ_i. A neuron i receives positive customers from the environment according to a Poisson process with rate λ_i^+ (no negative customers arrive from the environment).

The potential of a neuron is the number of positive customers in its waiting queue. When a neuron receives a positive customer, either from another neuron or from the environment, its potential is increased by 1. If a neuron with a strictly positive potential receives a negative customer its potential decreases by 1. After leaving neuron i, a customer either leaves the network with probability d_i, goes to queue j as a positive customer with probability P_{ij}^+ or as a negative customer with probability P_{ij}^-. So, if there are M neurons in the network, for all $i = 1,\ldots,M$:

$$d_i + \sum_{j=1}^{M} (p_{ij}^+ + p_{ij}^-) = 1$$

For sake of understanding, simplified description of how to use RNN as a learning tool is given here. A particular type of RNN (feedforwarding) is selected and trained with inputs. Knowing the values of a set of input (the mapping between inputs and outputs), RNN learns how to evaluate the function for any input. The function input is the vector $\vec{\lambda_i} = (\lambda_1, ..., \lambda_M)$ corresponding to M input parameters from the network. And the function output is the vector $\vec{\rho_i} = (\rho_1, ..., \rho_1)$, that means the stationary loads of the RNN. With this output vector, the RNN behaves similar to a function that gives MOS after entering argument values. The detailed description of RNN learning is out of the scope of this thesis, please refer to [17] for more information.

- **Methodology**
  In this section, description of how PSQA works is explained. Before using PSQA in real time, three steps need to be done beforehand. The whole process needs to be done for each given application.

  1. **Quality-affecting factors and Distorted Video Database Generation**
     In this first step, we select a set of quality affecting factors that have significant impact on quality such as codec, bandwidth, loss, delay, or jitter along with their ranges of values. A set of parameters with their specific values is called a *configuration*. A distorted video database is generated by varying the representative configurations. A set of quality affecting parameters (P parameters) such as codec, loss, delay, jitter... is written as $P = \{\pi_1, ..., \pi_P\}$, each has representative values with π_{min} and π_{max}. A set of values for each parameter is $\{p_{i1}, ..., p_{iHi}\}$ with $p_{i1} = \pi_{min}$ and $p_{iHi} = \pi_{max}$. For example, if a set of loss rate (in unit of %) is 0,1,2,5,10, then $H_i = 5, p_{i1} = 0, p_{i5} = 10$. A configuration Υ is $\{\upsilon_1, ..., \upsilon_p\}$ where υ_i represents one of chosen value for p_i. It can be noticed that the number of possible configurations ($\prod_{i=1..p} H_i$) is huge. Therefore, only a representative subset (S) needs to be selected for subjective evaluation; i.e. S configurations $\{\Upsilon_1, ..., \Upsilon_S\}$ where configuration $\Upsilon_S = \{\upsilon_{s1}, ..., \upsilon_{sP}\}$ and υ_{sP} is value of parameter π_P in configuration Υ_S. After that, M media samples are selected (σ_m with $m = 1, ..., M$). For each $\sigma_i, \{\sigma_{i1}, ..., \sigma_{iS}\}$ is a set of samples that have encountered S varied conditions when transmitted over the network. The implementation of this step could be done by experiments on real platform, network emulator, or network simulator.

  2. **Subjective Quality Assessment**
     In the second step, chosen configurations from the previous stage are evaluated by a subjective evaluation campaign. Single Stimulus method is used; a panel of human observers evaluates distorted videos as illustrated in Fig.3.4. Then, MOS is computed the same way as in subjective approach. Mappings of configurations and corresponding MOS are stored into two separated databases called *training* and *validation* database.

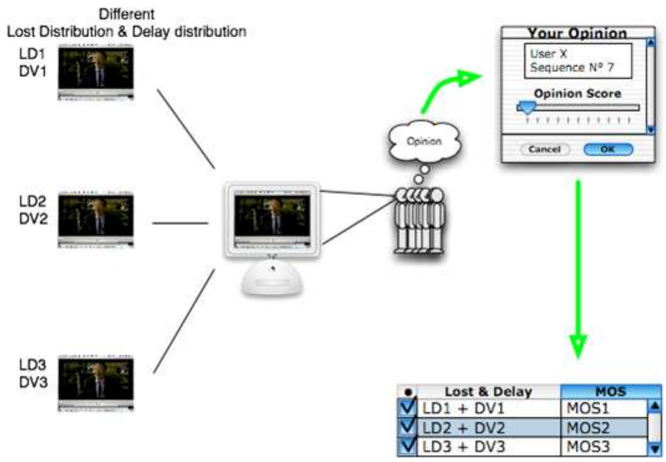
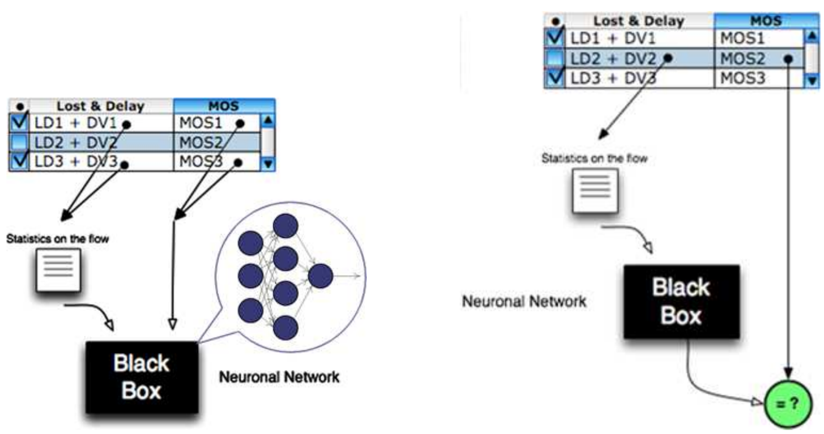


Figure 3.4: Subjective Quality Assessment Phase.



(a) Training (b) Validation

Figure 3.5: Learning of the quality behavior with RNN.

3. **Learning of the quality behavior with RNN**

   In the third step, the RNN learns the mapping of configurations and scores as defined in the training database. Once it has been trained, we get a function $f()$ that can map any possible value of parameters into MOS. The RNN is validated by comparing value given by this function $f()$ at the point corresponding to each configuration in the validation database (that RNN has not seen before). If the values are close enough, the RNN is validated. Otherwise, chosen configurations have to be reviewed and the step 1 through step 3 have to be repeated again until the RNN is validated. Once the RNN has been validated, PSQA is easy to use. To get instantaneous score at time t, we just measure quality-affecting parameters at time t and give them to RNN, which returns back the MOS value simultaneously. PSQA runs in real-time as if there were real humans marking their perception of quality.

   Once the RNN has been trained and validated, PSQA is very easy to use. It can be run anywhere in real time without any interaction with real humans. It is necessary to measure the quality-affecting (objective) parameters at time t and to evaluate these values with the RNN to obtain the instantaneous perceived quality at time t. PSQA gives scores in terms of MOS as if there were real humans marking their perception of quality.

3.3.3 Performance Evaluation

This section presents performance evaluation of the three methods. It begins with a description of scenario and environment of the test, implemented in the network simulator NS-2 [21]. Among several multimedia applications, investigation of performance is carried out for the video streaming application because it is one of the most popular applications today. In addition, further studies in this document will also concentrate on this application.

1. **Test Environment and Scenario**

   The interested environment is wireless networks IEEE 802.11 [80] standard since it is nowadays largely deployed. The network operates in infrastructure mode meaning that all traffic passes through an access point. The video sequence is an H.264-coded sequence (named "foreman") of duration 12 seconds and consists of 300 frames. It is encoded at 512 Kbps and streamed in unicast mode using UDP.

Loss Rate	0% 1% 2% 3% 4% 5% 6% 7% 8% 9% 10%
Mean Loss Burst Size	1 3 5

Table 3.3: Investigated configurations.

In the scenario, the client suffers from different loss rate and mean loss burst size indicated in Table 3.3. These loss rates ranging from 0% to 10% are chosen because it is interesting to see how QoE is affected with different loss percentage and how the three approaches behave. The rate does not go further than 10%, which is already a high loss rate; as higher rates will always provokes bad quality in the same way as 10% does. A simplified version of the Gilbert model [18] illustrated in Fig.3.6 is used for simulating burst of losses in the network. It is considered as a model that gives a good approximation of losses on the Internet. Parameters are: p, the probability of loss after a correct transmission, and q, the probability of correct transmission after a loss. The steady-state distribution of this model is given by $\pi_1 = q(p+q)^{-1}, \pi_0 = p(p+q)^{-1}$. The distribution of the length S of generic burst of losses, considering the system in equilibrium, is geometric: $Pr(S=n) = (1-q)^{n-1}q, n \geq 1$ with mean $E(S) = q^{-1}$. LR of the flow, according to this model, and the MLBS of the stream are: $LR = \frac{p}{p+q}, MLBS = E(S) = \frac{1}{q}$. Reciprocally, $q = \frac{1}{MLBS}, p = \frac{LR}{1-LR}\frac{1}{MLBS}$.

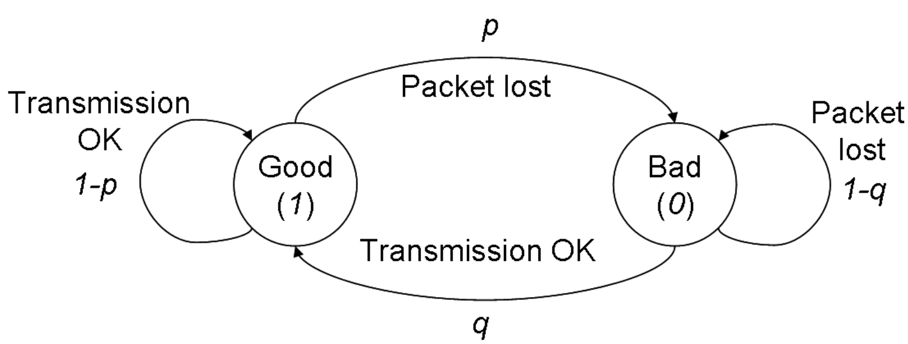


Figure 3.6: The simplified Gilbert model.

To select mean loss burst size (MLBS), some real experiments have been carried out to see distribution of occurrence of different burst size. D-ITG (Distributed Traffic Generator) [81] is used for varying load and QoSMet [82] for collecting statistics. Fig. 3.7 illustrates the proportion of mean loss burst size while varying load in the network from 0% to 80%. Seeing the partitioning of each size, investigations have been continued with three selected burst sizes (1,3, and 5) with respect to low, medium, and high burst size. However, it can be seen that the size of 1 has occurred the most often (76% of the time), hence the attention will be concentrated more on this size.

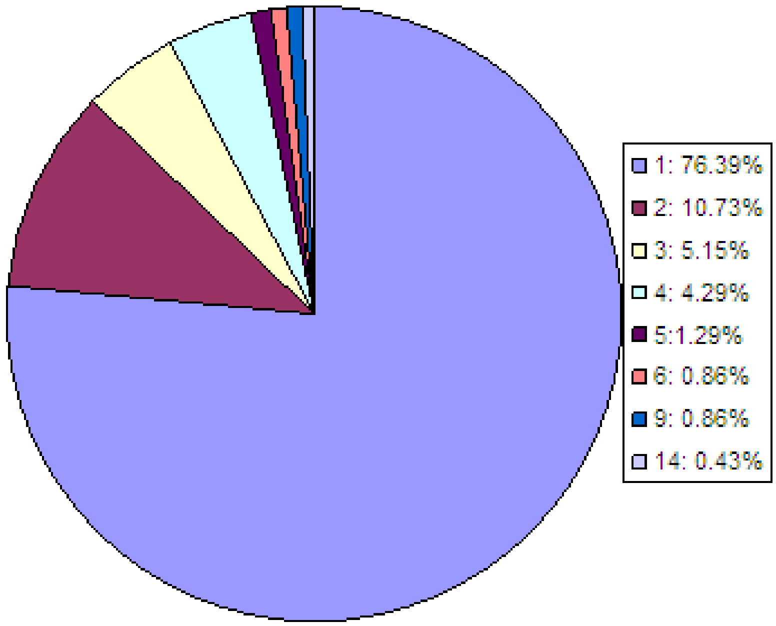


Figure 3.7: A proportion of loss burst size in real scenario.

2. **Implementation**
   Implementations are done with the network simulator NS-2 version 2.29 [21]. Wireless IEEE 802.11 implementation flaws of the original version are patched with wireless update patch from [83]. The patch includes realistic channel propagation, Ricean propagation model, 802.11 bug fixes, multiple data transmission rates support, and adaptive auto rate fallback (AARF). Implementation of video streaming application is done by adding a video packet transmission module (`videotrace`) in NS-2. This module enables transmission of parsed traces from real video sequences within NS-2. For communications between PSQA and NS2, PSQA module (`rnn`) has been integrated into this version of NS-2 so that it can directly acquire input statistics for RNN. The development of this module is based on RNN source code from colleagues in DIONYSOS research group [84]. The basic code contains all functionalities necessary for using RNN such as creation, training, and validation. The interactions between RNN and NS-2 have been implemented in order to enable communications of RNN input/output with NS-2. For getting score, PSQA is called every *t* second (here *t=1*).

   To get PSNR and subjective evaluation, the procedure is illustrated in Fig. 3.8. First of all, a raw YUV digital video sequence is processed by an encoder that generates the H.264 bitstream. The bitstream is then parsed to get a trace file compatible with NS-2 network simulator. For each run, the simulation is specified with desired loss rate and MLBS (using simplified Gilbert model). As a result, loss trace file is obtained, which is used by the loss insertion block (`gilbertloss`) to erase lost packets from the original H.264 bitstream. Finally the distorted bitstream is decoded to a raw video file for visualization and quality assessment as well as for PSNR computation. For subjective evaluation, observers are asked for the average impression of each video sequence. Single Stimulus with five expert observers is used and then an average score are computed to represent MOS.

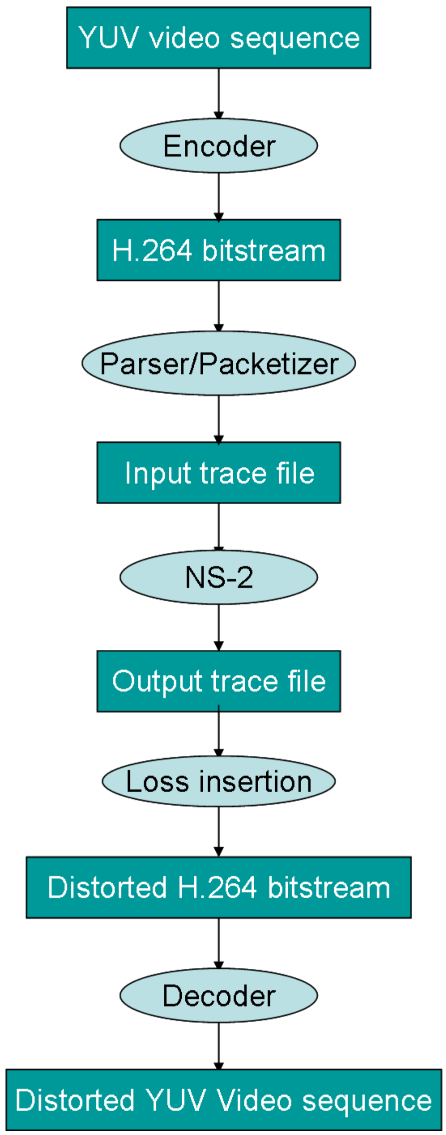


Figure 3.8: Process of trace file generation and video distortion.

3. **Results**

The results given by subjective evaluation, PSQA, and PSNR are presented respectively. Each of them is discussed and a comparison is done by using subjective evaluation as reference since the goal here is to approximate QoE as accurately as possible comparing to the scores obtained with subjective approach.

- *Subjective Scores*

  The Fig. 3.9 shows the scores along with standard deviations obtained with subjective evaluation. It can be seen clearly that loss rate has a great impact on perceptual quality as we can see in this figure the degradation in quality increases while loss rate increases. However, the degradation is not proportional to loss and it does not represent any function of loss rate. That is why only technical parameter cannot reveal subjective quality. We can observe that loss burst size also has an impact on the end-user perception especially with this video application. We can see that the highest burst size 5 results in better quality than other two lower sizes in almost every loss rate except with very high loss from ranging 8% to 9%. This can be explained by the fact that human observers prefer to have small number of grouped loss than high number of dispersed loss. However, for very high rate (8%, 9%, 10%), the degradation is too important that the quality is no longer acceptable in all cases anyway.

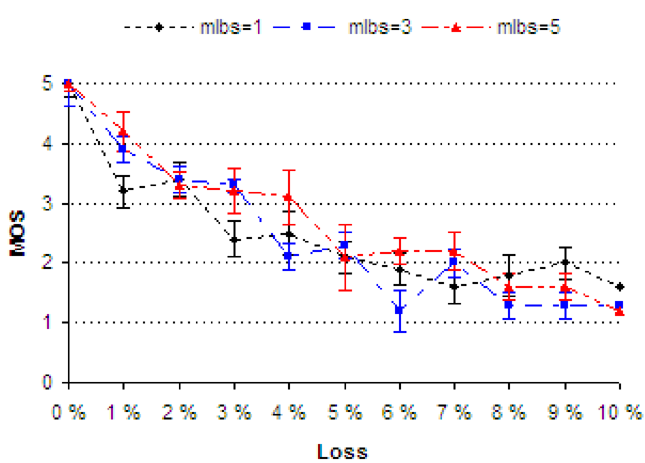
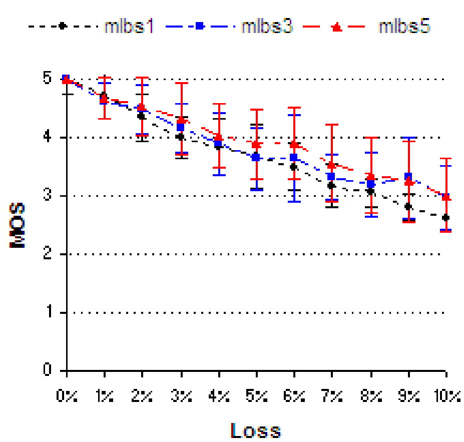


Figure 3.9: MOS obtained with subjective approach.

- *PSQA Scores*

  Fig. 3.10 (*left*) shows the PSQA scores along with standard deviations for videos experimenting different loss rates and mean loss burst sizes. Each point is computed by taking an average PSQA score of the whole duration (12 seconds). We observed that better QoE is obtained with higher MLBS, similar to what we have seen in Fig.3.9 with subjective evaluation. However, we also observe that PSQA scores are too high comparing to the QoE really obtained by subjective evaluation, so investigation has been conducted and it is found that if we take minimum score for each video (cf. Fig.3.10 *right*), we get better precision of quality as we will see later in comparison section that the graph follows the one of subjective evaluation better than the average score. Thus, in the following the comparison continues using the minimum value instead of the average value. The reason why minimum scores show better precision is because humans always pay more intention to the period of video where they have seen the worst quality, this event is more remarkable than the no-loss period. Therefore, they give quite pessimistic scores than what they have really seen in overall duration of the sequence. For more details on how scores evaluate in time during the play-out, individual scores for the three different loss rates are also shown in Fig.3.14: 2% for low loss, 5% for medium loss, and 10% for high loss.



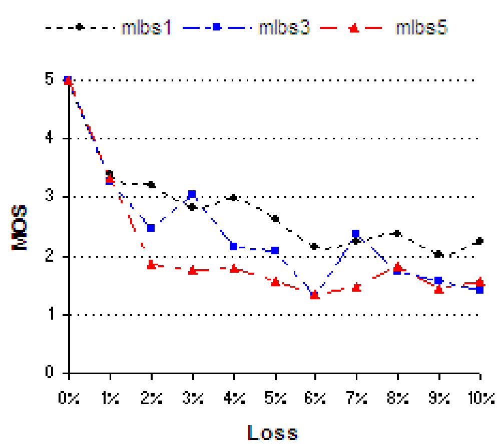


Figure 3.10: MOS obtained with PSQA: Average (*left*) and Minimum (*right*).

- *PSNR Scores*

  First, overall PSNR of the video (meaning the average PSNR of 300 frames) is computed and it is then converted to MOS according to Table 3.2. Fig.3.11 (*left*) shows this overall PSNR for each configuration. Fig.3.15 shows three PSNR graphs of individual frame corresponding to three different loss rates: 2% for low loss, 5% for medium loss, and 10% for high loss; it illustrates more in details how PSNR of each frame evaluates during the video duration.

  The similar situation as in PSQA happened in PSNR, the average PSNR of each frame gives very bad approximation that does not correlate well to subjective evaluation. The investigation is done and it is found that we should better use the average MSE of the overall frames and compute PSNR with this value (cf. Fig.3.11 *right*). Better approximation of quality is obtained as we will see later in comparison section; the graph follows the one of subjective evaluation far better than the average PSNR of each frame.

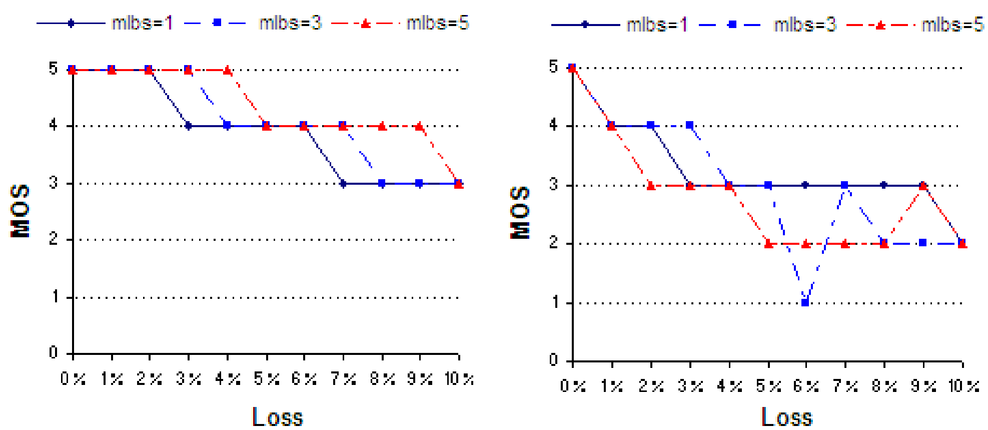


Figure 3.11: MOS obtained with PSNR: Average of PSNR (*left*) and PSNR of average MSE (*right*).

- *Score Comparison*

  Fig.3.12 and Fig.3.13 show comparison between PSQA (hybrid approach) and PSNR (objective approach) with reference to Single Stimulus (subjective approach). We can see that PSQA outperforms PSNR in case of MLBS=1 and MLBS=3. However, PSQA performs worst in case of MLBS=5. This can be explained by the fact that when minimum score of the whole duration is considered, in case of high burst, PSQA will give quite bad score as we can see in Fig.3.14 that minimums of all graphs with MLBS=5 are very low. Nevertheless, this situation of high burst size happens rarely in real scenario (1% according to Fig.3.7). Even though in high burst size

PSQA performs worst than PSNR but its advantage of accurate real-time measurement in other cases, contrary to post measurement of PSNR, makes it more attractive for resource management mechanisms.

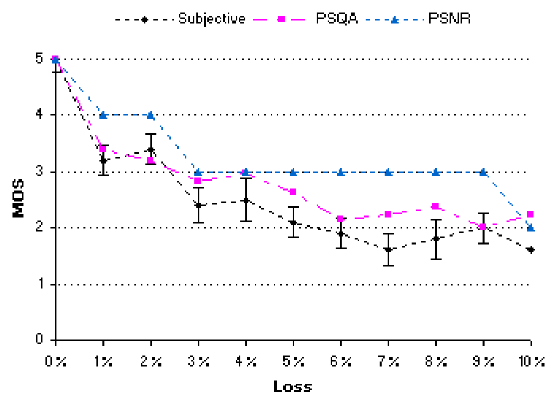
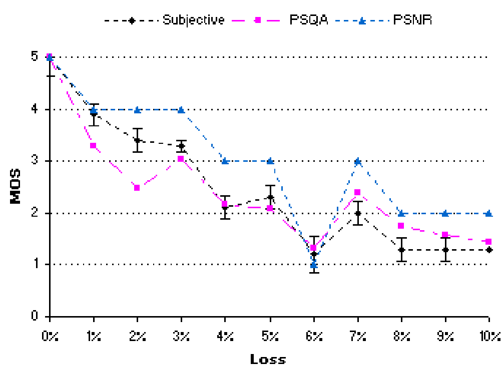
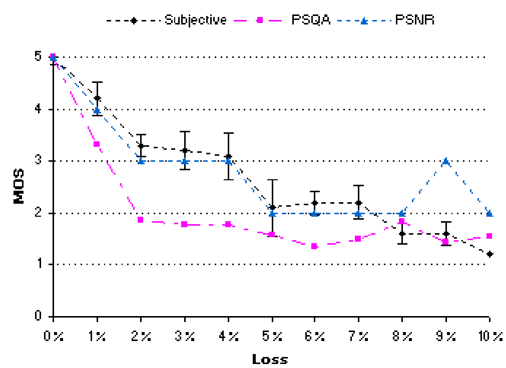


Figure 3.12: Comparison of MOS obtained with the 3 approaches: MLBS=1.



(a) MLBS=3.

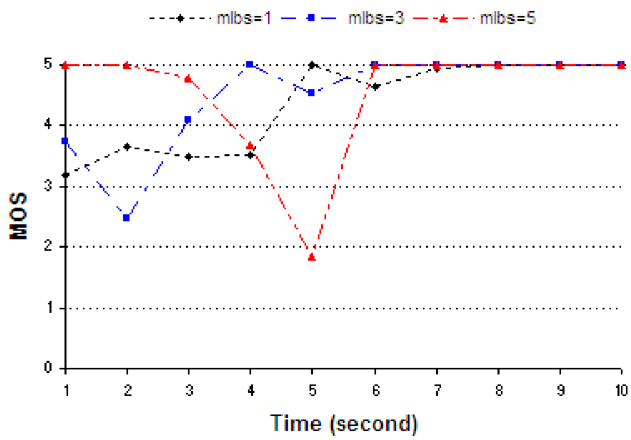


(b) MLBS=5.

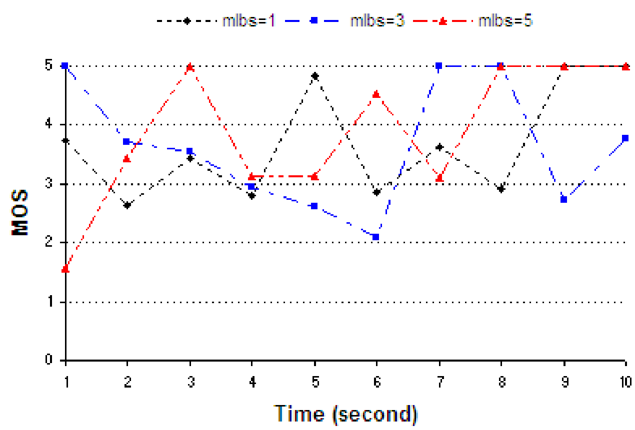
Figure 3.13: Comparison of MOS obtained with the 3 approaches.

Moreover, it can be noticed from Fig. 3.15 that PSNR scores fluctuate in times and PSQA scores are more stable (Fig. 3.14). As such, the latter is better to use for adaptation mechanism since if we use PSNR, which changes often, the mechanism will have to adapt often and may result in instability of the system.

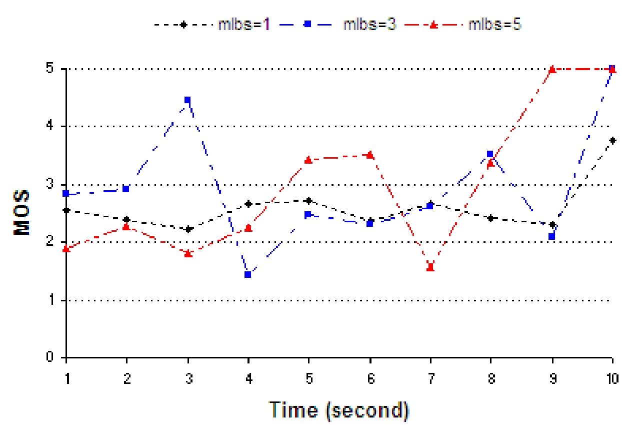
Finally, the main contribution of this subsection is the performance evaluation of PSQA for video streaming application in WLANs. After comparing PSQA to PSNR and subjective approaches, usage of QoE as metric for resource management is validated and thus discussion about opportunities enabled by QoE metric (via PSQA) can be presented in the following section.



(a) Low loss (2%).

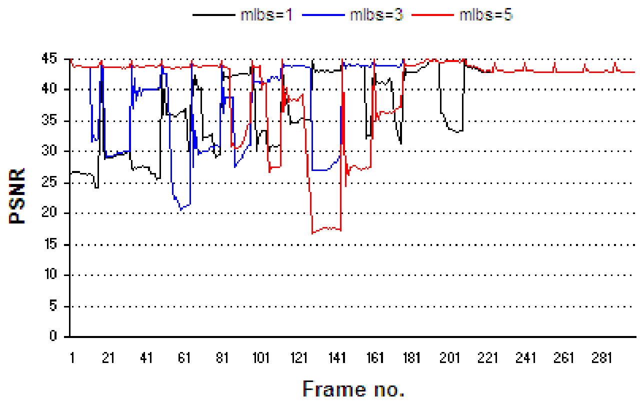


(b) Medium loss (5%).

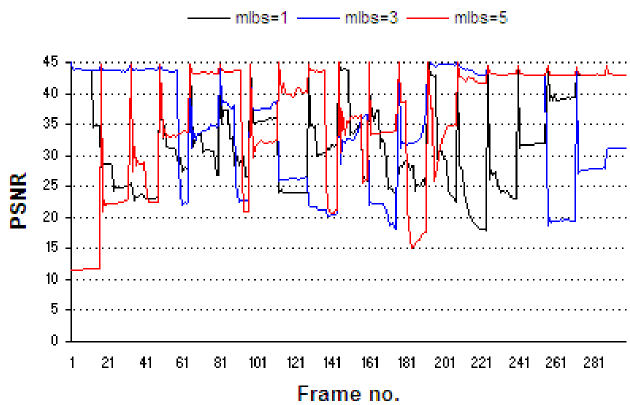


(c) High loss (10%).

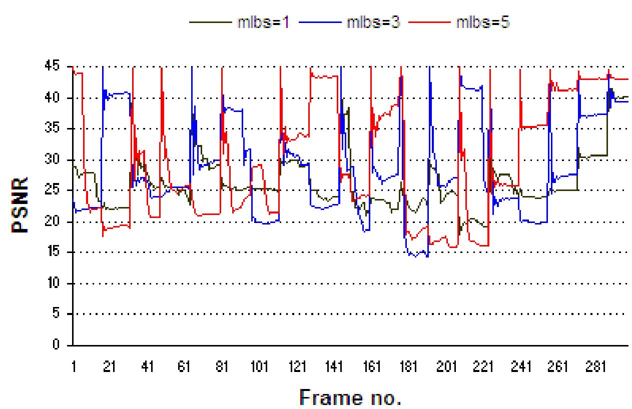
Figure 3.14: MOS of different loss rate obtained with PSQA.



(a) Low loss (2%).



(b) Medium loss (5%).



(c) High loss (10%).

Figure 3.15: PSNR of different loss rate.

3.4 Applying QoE for Resource Management

We have seen that PSQA gains advantages and avoids drawbacks from both subjective and objective approaches. It is accurate and can be run in real time; moreover, it is less time-consuming and it does not require manpower (except in the subjective quality assessment step, which is normally done only once). Having capability of assessing QoE automatically via PSQA opens a wide range of opportunities. QoE then becomes an interesting metric for managing network resources. Since competition between network operators will be based on this metric (i.e. clients all want to have the best perception of their multimedia applications), it is important to explore management directions enabled by QoE concept.

Some previous works have been conducted using this PSQA technique, for example, VoIP over wireless LANs [85], video application over DiffServ networks, or IPTV over peer-to-peer networks [86]. However, PSQA is application-dependent and system-dependent, hence even previous related works have been done but none considers video streaming over WLANs, which is becoming very popular nowadays. In addition, main objective of previous works is often in quality monitoring purpose whereas objective of this dissertation is to use QoE for network management purpose.

3.4.1 Applying PSQA in Resource Monitoring

The QoE monitoring can be done at different levels, either at end-user or at network, depending on the purpose. The advantage of measuring at end-user is accuracy because monitoring entity is located at user itself and real-time information can be collected easily. On the other hand, monitoring entity can also be placed at network side. This means PSQA can be run at router or point of attachment (PoA) for being able to react directly to current situation.

Fig. 3.16 illustrates different levels in the network where PSQA can be placed; it also gives an example of monitoring at end user. It can be noticed that monitoring at end user will give the most accurate information. This information can be either used at user terminal (adaptation at end-system) or it can be forwarded to access point level where first set of solutions can be executed by the access network. If collected information is not enough, access points can forward their information to access router where more local data is available. Finally, if global decision needs to be made then access routers can, in turn, forward collected information to central controllers who can make decision with global view of the system.

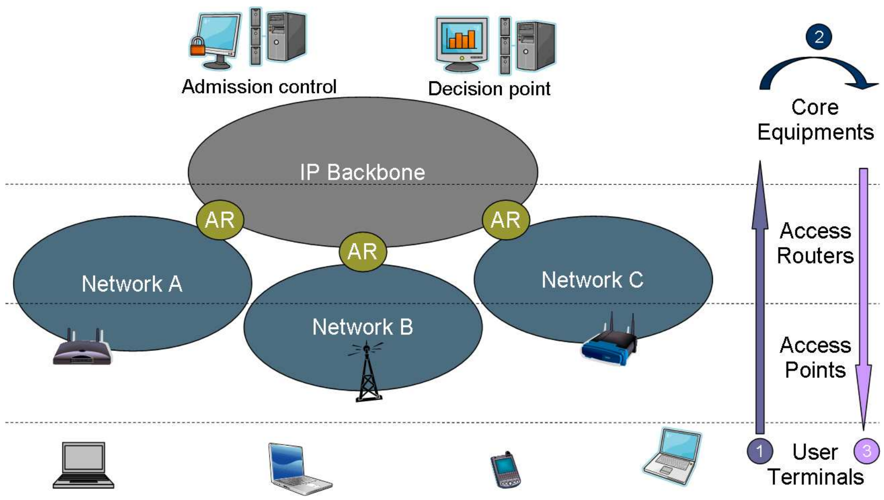


Figure 3.16: Different levels of control.

The following gives a more detailed example of how to use PSQA in QoE monitoring and also how communications between network entities can be processed:

- **PSQA running at end-user side**

  Fig. 3.17 depicts communications between PSQA, user terminal, and access point; in which PSQA is placed at user level. For this case, user terminal directly feeds inputs concerning statistics of the flow to PSQA who computes MOS simultaneously and returns it to user terminal afterwards (1). Access point can then inform other users about the current QoE of ongoing users (2). An incoming terminal hence has valuable information about network performance (3).

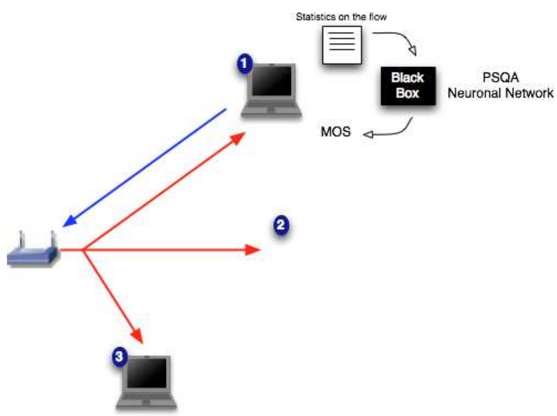


Figure 3.17: PSQA running at end user.

- **PSQA running at network side**

  Fig. 3.18 depicts communications between PSQA and access point, in which PSQA is placed at access point. This case is feasible when all necessary statistics concerning users can be collected at access point level. In this case, access point directly feeds inputs concerning statistics of the flow to PSQA who computes MOS simultaneously and returns it afterwards. Access point can then inform other users about the current QoE of ongoing users or manage network resource internally.

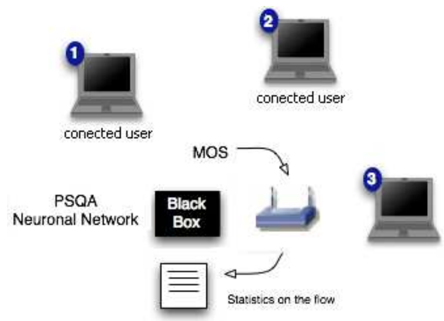


Figure 3.18: PSQA running at access point.

3.4.2 Examples of use case

This section provides some examples of use case that deploy QoE for resource management. Furthermore, it also gives directions about QoE for network management in terms of global system.

- *Call admission control mechanism*
  Increasing number of wireless users has pushed network operator to consider call admission control during operation. With information such as QoE of ongoing users, it is more efficient to perform the control than using only bandwidth- or throughput-related information. The experiencing quality of already connected users can imply the quality of incoming user; however, degradation after the admission has to be considered and handled properly. This can be done at different level, for example, PSQA can run at the access point in order to monitor quality of experience of ongoing users and filter the access from new users accordingly.

- *Bandwidth allocation*
  QoE can help network operator handling resource allocation. It would be advantageous to regulate bandwidth taking into account user experience. For example, in variable bit rate (VBR) traffic, bit rate varies often resulting in different amounts of bandwidth requirement in specific times. At those moments when the need of bandwidth is less, operators can give the remaining to other traffics. Hence, they can earn more benefit while keeping customers satisfied.

- *Network selection*
  Increasing number of network operators has also pushed users to select the network that will provide the highest quality. This can be done, for example, at the access point level with the help of communication protocol such as IEEE 802.11k [80]. Access points can broadcast information about its own network using one of the specified report frames.

- *Handover management*
  When mobile users move from one location to another, it is possible that network operator guides user to connect to a specific network according to QoE from ongoing connections in candidate cells. Some works [69] have already been done to provide a method for execution of this guidance. The combination of the two will result in more quality-related ways of control.

- *Heterogeneous network management*
  Moreover, seeing that QoE is independent of network technology, we can imagine using QoE as metric in other network environments as well. So far, some works have been done in different networks such as Peer-to-Peer [87]; however, the authors focused more on network monitoring aspect. Thus, the management issue is still left for investigations. Moreover, a growing heterogeneous network can also be another target. QoE is a perfect metric in this type of environment where different network technologies co-exist with the same goal of providing the best service to users.

- *QoE-aware Framework*
  Finally, in the future we can also imagine a QoE-oriented framework, a whole network infrastructure based on quality of experience. For communications between each network entity, more studies are needed. However, there exist already some supports provided by the standard such as IEEE 802.21 [31], which provides tools to handle handover execution between different technologies. For SLA, different levels of user would be established along with differentiated service quality and price. Fig.3.19 depicts a possible QoE-aware framework.

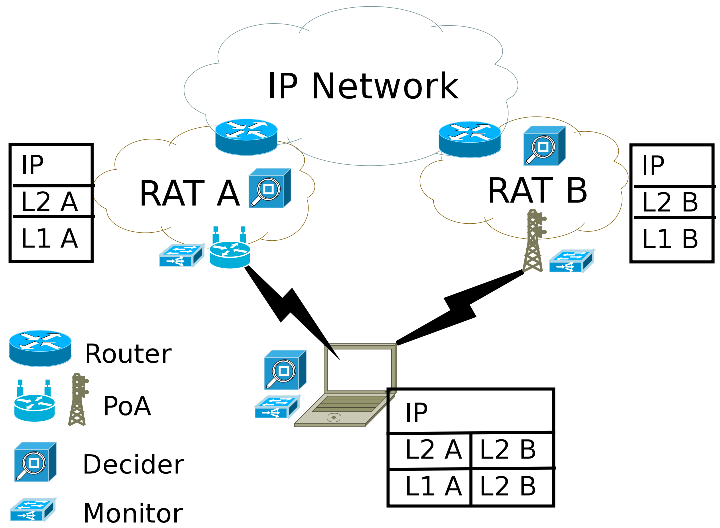


Figure 3.19: QoE in Resource Management.

3.5 Conclusions

The concept of quality of experience (QoE) has been introduced in this chapter. Different QoE measurement approaches are described and performance evaluation has been conducted. PSQA (Pseudo-Subjective Quality Assessment) is evaluated and its performance demonstrates that it represents an appropriate way to access user experience in real-time manner. After we have seen that it is possible to automatically evaluate QoE using PSQA, we foresee resource management from a different angle.

Therefore, reader is introduced to a novel resource management approach using quality of experience as metric in many management mechanisms and framework. It is more relevant and more flexible than QoS parameters when dealing with multimedia applications. In the following parts of this dissertation, deeper investigations on using QoE in network management will be conducted. Use cases will be studied both from network and user perspective and within both homogeneous and heterogeneous environment.

Part II

Network-centric Resource Management

We have seen in Part I state of the art on radio resource management and importance of user experience in wireless multimedia networking. In this Part II, deeper investigations on using QoE in network management will be conducted from network operator perspective or what we call network-centric approach in this document. As we can see that wireless networks nowadays raise many problems for network operators to manage their resources. These problems come mainly from restricted bandwidth and variable radio condition inherited from the wireless nature of network. In addition, the emergence of multimedia traffic with its strict requirements make it more complicated to manage network and hence efficient management mechanisms are indispensable. In this part, three common mechanisms, namely admission control, rate adaptation, and scheduling, will be discussed. For each mechanism, a scheme based on user experience will be presented. The human QoE is obtained by PSQA tool previously described. Instead of relying on technical parameters such as bandwidth, loss, or latency, which do not correlate well with human perception, this part of dissertation demonstrates how QoE can be used as metric in network management. The simulations have been done in wireless IEEE 802.11 and mobile UMTS environments respectively.

Chapter 4

Admission Control

4.1 Introduction

Since wireless local area networks have started to be deployed, users can connect easily to the Internet and the number of Internet users has increased significantly. Nowadays, WLAN based on IEEE 802.11 [80] standards with infrastructure mode is the most popular as we can see hotspots everywhere. At the same time, enormous progress has been made with this technology, and the ability to support advanced services became possible. As a result, mobile hosts running real-time multimedia applications such as video streaming and VoIP are ubiquitous. These multimedia users are the major concern because their traffic is restricted in terms of quality. In addition, the nature of wireless network (limited bandwidth, shared resources, channel interference...) made it easy to be over-utilized. Consequently, network load must be controlled carefully so that acceptable quality for real-time applications can be maintained while not penalizing network operators with underutilization.

Therefore, this dissertation firstly presents an *admission control* mechanism based on quality of experience perceived by ongoing users. The proposed scheme is based on Mean Opinion Score (MOS) and without interaction from real humans (via PSQA). The simulation results will demonstrate the performance of this approach compared to the loss-based approach regarding user satisfaction evaluated by the QoE achieved at user and bandwidth utilization of the network evaluated by the goodput.

The remainder of this chapter is organized as follows. The chapter begins with related works in Section 4.2. Then, Section 4.3 presents the proposed admission control algorithm along with interaction between access points and PSQA tool. Implementations are described in Section 4.4 and the corresponding simulation results are presented in Section 4.5. Finally, Section 4.6 presents conclusions.

4.2 Related Works

To guarantee service quality at users and to optimize resource utilization, admission control is essential; otherwise degradation in quality will result from high collision. Controlling admission can be handled with several methods, we can observe two main approaches: *access scheduling* and *resource provisioning*, as explained below.

- The first approach consists in scheduling access to the wireless channel. This approach has been proposed to solve inherited problems from Media Access Control (MAC) protocol in IEEE 802.11 standard that does not support neither quality of service nor multiple traffic categories. First, IEEE 802.11e [88] standard has been created for supporting multiple traffic categories and then many variations have been designed, most of them try to schedule access to the channel taking into account different traffic categories and prioritizing multimedia traffic. Similar approach proposed in [89] manages resources by splitting the contention period into two subperiods: one for contention between real-time stations and another for contention between non real-time stations.

- The second approach consists in restricting the volume of traffic that enters into the network with an objective of QoS provisioning. This is usually done by estimating channel utilization based on network measurements. Some schemes have focused on the analysis of throughput in saturated conditions; referring to collision probability analyzed in [90], the author of [91] provide a mechanism to predict achievable throughput for all users after a new connection is accepted. Another scheme proposed by in [92] has developed an analytical model to assess the capability of 802.11 and to control admission of new flows based on channel busyness ratio.

Even many admission control mechanisms exist, most of them are only aware of QoS and very few takes into consideration the quality of experience, which is the most important factor in the increasing multimedia traffic today. A majority of the mechanisms rely on technical parameters, especially bandwidth. They usually compare available and requested bandwidth before deciding whether to accept a connection or not, similarly to resource provisioning approach. This works well with wired networks where bandwidth provisioning is easier (due to stable network condition) than in the wireless environment. In addition, bandwidth alone is not enough to guarantee quality. To accomplish both goals of enabling high quality for admitted flows this chapter proposes a QoE-based admission control mechanism that administers the access network in real time based on user's perceived quality.

4.3 QoE-based Admission Control Mechanism

This section presents an admission control mechanism based on user experience with help of PSQA tool. The interested context is wireless access environments such as IEEE 802.11 standards with infrastructure mode (all traffic passes through the access point). This choice has been made because the access point can act as controller equipped with PSQA tool. The main idea is to have access points monitor MOS of its ongoing connections in order to have knowledge about the perceived quality level of the service and then to decide whether to accept a new connection or not accordingly.

4.3.1 Access points in the proposed scheme

Access point (AP) in the admission control algorithm can be illustrated with a Mealy automaton in Fig.4.1; focus is only on the states concerning the proposed scheme. Assume at the beginning that access point is up and waiting for connection requests. When a new connection is requested, the access point computes an average MOS of all ongoing connections. If MOS is higher than an acceptable level plus a threshold, then a new connection will be accepted (AP returns then to initial state); otherwise it will be rejected (AP then waits until connection release before returning to initial state). The threshold is used to absorb degradation of quality after a new connection is accepted. For economizing processing time, the access point computes MOS only when a new connection is requested and not periodically. For example, an access point computes MOS after receiving an *Association Request* from a station then it sends back *Association Reply* according to the algorithm. This approach is dynamic and economic thus it is suitable for wireless networks where channel conditions change often.

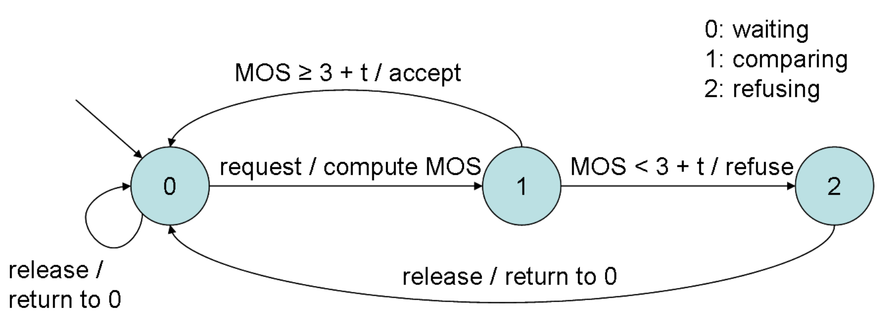


Figure 4.1: States of the access point in the proposed scheme.

In the mechanism, the score 3 (*fair* quality), according to 5-point scale in Fig. 3.1, is selected to decide for admission as it is known that this quality level is acceptable for video streaming applications. It can be noticed that the threshold t is very delicate to define as it depends on the granularity expected. If t is high, it will result in high quality because the scheme will grant all network capacity to a small number of flows.

Nevertheless, this restriction may raise under-utilization problem, which is expensive for network operator. With the similar reasoning, if t is small, it will result in low quality due to congestion in the network. Thus, a tradeoff between bandwidth utilization (accepting more connections) and its consequence in connection degradation has to be weighed properly.

4.3.2 Interaction between access points and PSQA

All access points in the scheme are assumed having two additional functionalities: monitoring loss and communicating with PSQA. The PSQA tool operates at every access point and helps them for MOS computation. The interaction between access point and PSQA tool is explained and illustrate in Fig. 4.2.

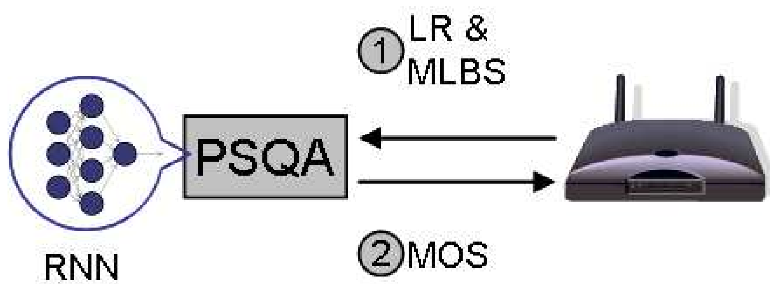


Figure 4.2: Interaction between the access point and PSQA tool.

1. The access point monitors loss statistics of each caring station and gives these statistics to PSQA tool as input.

2. After receiving statistics from the access point, PSQA tool computes MOS and returns it to the access point afterwards.

Two specific parameters concerning losses are monitored because previous work of PSQA has demonstrated that loss statistics is the most important factor for quality. Therefore, the statistics considered in the implementation are loss rate (LR), the loss rate of video packets; and mean loss burst size (MLBS), the average length of sequences of consecutive lost packets, this parameter captures the way losses are distributed in the flow as it affects dramatically the perception of video [86]. High MLBS makes impairment more visible in the video; however, after the study in [93], it has been found that humans prefer sequences where losses are concentrated over those where losses are isolated but more frequent.

4.3.3 Example of scenario

This section first illustrates the effect of no admission control in the system and then it explains how the scheme will be applied to this type of situation.

Assuming in this example that the network operator does not implement any admission control mechanism; the connection arrival rate is one connection per second and the network operator accepts all connections until its maximum capacity. With simulations, we observe how quality changes in time. From the Fig.4.3, we can see that QoE is excellent at the beginning because a small number of connections can profit from all available bandwidth. However, when the number of admitted connections gets to 11, the quality begins to degrade until it reaches and remains at the score 1 (*bad* quality).

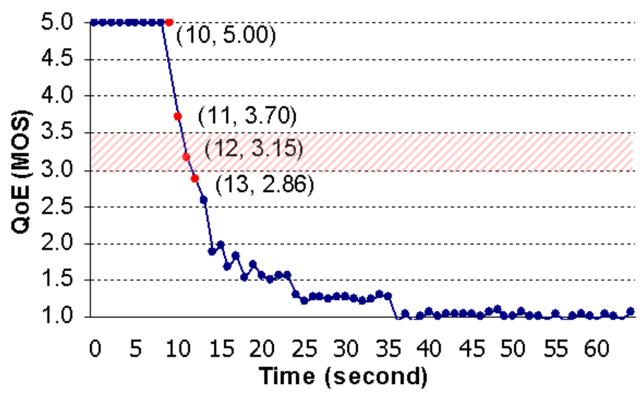


Figure 4.3: QoE in the example scenario.

The functioning of the scheme is illustrated by this example. In this scenario, a value of 0.5 is adopted for the threshold t because the chosen value provides a good balance between bandwidth utilization and quality degradation, after extensive simulations. With $t = 0.5$, the network operator will accept connections until current MOS reaches an interval [3.0-3.5] and then stop admitting new connection at 12 admitted flows where this threshold is attained (hatched zone). The number of flows remains 12 until at least one connection releases its bandwidth and the operator can then accept a new flow again.

It can be noticed that the threshold t has a great impact on the number of admitting connection. In the future, different thresholds could be used to treat different user priorities. For example, user with high priority will have a high threshold because the perceived quality have to be guaranteed strictly. On the other hands, the threshold for lower priority users may be smaller because this class of users is less sensitive or less restricted in terms of quality. However, some cautions should set up so that the small threshold of low-priority user should not affect quality of already admitted users.

4.4 Performance Evaluation

The performance evaluation is conducted in wireless access network based on IEEE 802.11b specification [94] and the scheme is evaluated with video streaming application. For the test, the network simulator NS-2 [21] version 2.28 is used with the wireless update patch from TKN [95]. Two extra modules (`videotrace` and `rnn`) previously explained in Chapter 3 are integrated into the simulator and admission control is done according to the described algorithm.

4.4.1 Simulation setup

The admission control is implemented in access point operating on infrastructure mode. For the topology, the access point is situated in the middle of the cell possessing a coverage area of 500mx500m. Client nodes are positioned randomly in the cell. Each client requests for video streaming of 384 kbps with connection handling time of 60 seconds. Connection arrival rate is one connection per second. The access point monitors user experience for each connection with PSQA to compute MOS of each connection using statistics measured at its downlink interfaces. The scenario is similar to the one explained in the example scenario.

4.4.2 Comparison with loss rate based schemes

The proposed scheme is compared with admission control implementation based on loss rate because it is the pertinent metric that is widely used to determining service quality. Three rate-based schemes (2%, 5%, and 10%) have been chosen, they correspond to low, medium, and high loss for video streaming application in the wireless system. In each loss-based scheme, the access point will stop admitting new connection when the specified percentage of loss is reached. The result of evaluation is detailed in the following section.

4.5 Results

Two significant metrics are considered for evaluation of the proposed scheme. The first one is user satisfaction that can be measured in terms of Mean Opinion Score (MOS) and the second one is bandwidth utilization that can be measured in terms of goodput. The results are explained according to these metrics and the summary of performance comparison is given at the end.

4.5.1 User satisfaction

For measuring user satisfaction, we should evaluate how users perceive the service and how satisfied they are. To do so, PSQA tool is deployed to measure MOS of each connection; this MOS expresses user perception of the service. The global MOS of the system is illustrated here by taking an average MOS of all active connections; this is done every decision epoch determined by the connection arrival rate. In this scenario, it is done every second, thus the MOS presented in Fig.4.4 is taken every second. In addition, PSQA also continues to measure MOS during transmission periodically in order to see how quality evaluates in time.

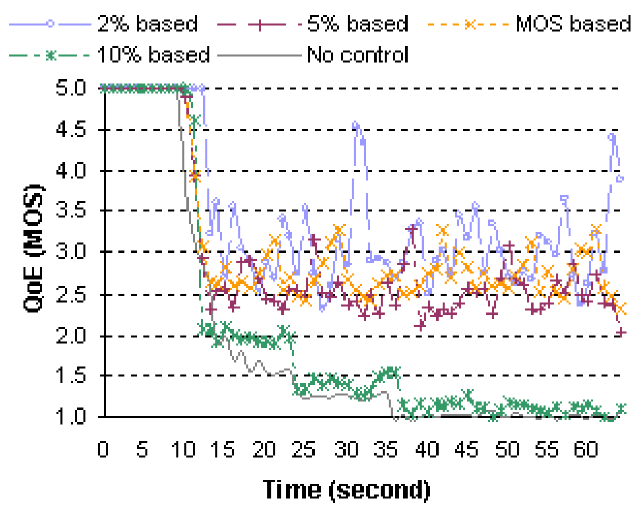


Figure 4.4: QoE and loss based schemes: MOS comparison.

From the Fig.4.4, it can be seen that the proposed scheme outperforms no-control approach and 10% based approach. This can be explained by the fact that in the no-control approach, the call admission control does not exist and the network always accepts new connections leading to congestion and hence bad quality. In the second case, limiting loss rate at 10% is too high for obtaining good quality for video streaming applications. The proposed scheme performs slightly better than 5% based approach which is, generally, a delimited loss rate beyond which quality will no longer be acceptable. Nevertheless, 2% based approach gives better scores than the proposed scheme does but with the price of bandwidth underutilization discussed in next section.

4.5.2 Goodput

For resource utilization assessment, the goodput is measured. It is, indeed, the application level throughput. It represents the number of useful bits per unit of time forwarded by the network from a source to a destination. For measuring the goodput

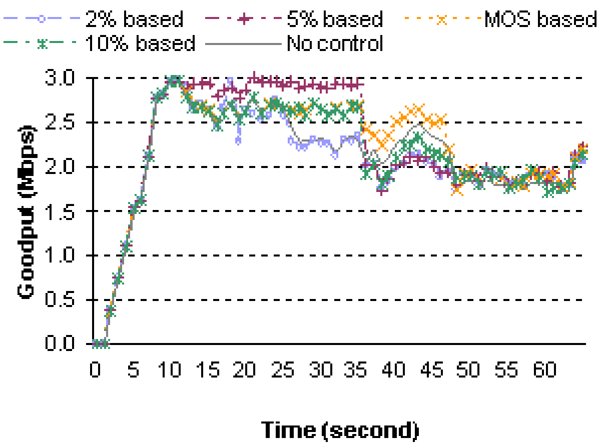


Figure 4.5: QoE and loss based schemes: Goodput comparison.

in NS-2, the number of bits successfully received at each station is computed. Fig. 4.5 shows the global goodput of the network in each scheme. The result confirms less goodput obtained by 2% based scheme as just mentioned earlier. In fact, the loss of 2% is very restricted for admission; consequently the goodput of 2% based scheme is lower than others. We can also observe that 5% based scheme perform pretty well in the beginning but the throughput drops sharply around the 35^{th} second. On the contrary, the proposed scheme performs slightly lower at the beginning but it maintains at good level until around 50^{th} second, while the others perform worst.

4.5.3 Performance summary

Table 4.1 summarizes the performance of five schemes previously explained and highlight the performance of the proposed mechanism. It also gives information about number of flows admitted by each scheme and the maximum bandwidth utilization.

Table 4.1: Comparing QoE and loss based schemes: Summary.

Scheme	Max. Bandwidth utilization	Connection admitted	Average MOS
2% based	3.6 Mbps	10 flows	3.62
MOS based	**4.32 Mbps**	**12 flows**	**3.35**
5% based	3.96 Mbps	11 flows	3.19
10% based	4.68 Mbps	13 flows	2.17
No-control	7.2 Mbps	20 flows	2.06

4.6 Conclusions and Perspectives

In this chapter, an admission control mechanism based on perceptual quality has been proposed. While others consider purely technical parameters, this chapter investigates the interpretation of these parameters to human perception (i.e. QoE). Thus, the proposed scheme provides a method to control radio resource while being aware of quality experienced by users.

In fact, the access point in the proposed scheme monitors quality experienced by ongoing users (with help of PSQA) and makes admission decision accordingly. Although the scheme is based only on current score (perceived by active users); we can obtain good performances. Furthermore, it would be interesting to see if we can improve the scheme to enable MOS prediction and hence *QoE provisioning*. If prediction of QoE is possible, operator would be able to manage resources with precision of QoE that will be reached by users.

Moreover, as service differentiation is an important concern in wireless LAN nowadays, it would also be interesting to study different thresholds to be used further to address different user priorities or traffic classes. Also, as admission control only solve the problem at network entrance phase, network condition may change during connection holding time, and thus adaptation along connection duration would also be another attractive issue for investigation.

Chapter 5

Multicast Rate Adaptation

5.1 Introduction

We have seen in the previous chapter how QoE can be used for admission in WLAN, this chapter will investigate multicast transmission and present how QoE can also be used to improve quality performance in this type of network. As we can notice, wireless networks have been deploying everywhere with IEEE 802.11 as the most popular standard; however, wireless resources are scarce and wireless condition varies often as mentioned previously. These limitations are crucial for applications with tight QoS requirements such as video or voice over IP. To cope with these problems, the standard has provided many features including *multi-rate* capability, which is the focus of this chapter. Multi-rate capability is beneficial especially for multicast transmission, in which the traffic sent by default at basic rate may result in capacity wasting due to longer channel occupancy. Moreover, the lack of feedback mechanism makes it difficult to deal with reliability or service quality. Some protocols have been proposed to use rate adaptation in handling the problem but none of them takes into account user experience, which is an essential quality indicator for multimedia application.

Therefore, in this chapter, two *rate-adaptation* mechanisms based on quality of experience will be proposed. The first one simply uses QoE as indicator to adapt transmission rate. Similar protocols have been proposed in the literature and most of them make use of the same static-threshold approach in order to decide when to change transmission rate. Unfortunately, static threshold does not adapt well to varying network condition, which is common in wireless environment. Thus, the second mechanism is also proposed; it provides dynamic rate adaptation mechanism based on quality of experience as well. For both schemes, PSQA tool is used for obtaining mean opinion score in real time. The objective is to improve bandwidth utilization while satisfying user experience. The results illustrate significant performance improvement obtained by the proposed scheme according to this goal.

The rest of this chapter is organized as follow. This chapter begins with backgrounds on wireless multicast in Section 5.2 following by related works in Section 5.3. The proposed schemes are presented in Section 5.4 along with their simulation set up and results. Finally, conclusions are given in Section 5.5.

5.2 Wireless Multicast

Multicast over wireless networks is a fundamental communication function because wireless network is inherently broadcast by nature. This means that a packet that is sent only once, will reach all intended recipients in multicast group. Therefore, multicast is an efficient method to transmit the same data to a group as it allows transmission of data to multiple destinations using fewer network resources. More recently, the fast-growth of wireless network and its application has pushed the deployment of multicast communication over wireless networks. We can see today, various applications support multicast; for example, conference meeting, mobile commerce (mobile auctions), military command and control, distance education, entertainment service, and intelligent transport system.

However, multicast application has some constraints. Multicast traffic has been set to the lowest transmission rate (basic rate) in order to reach all mobile nodes especially the further ones because they are subject to important signal fading and interference. The lower rates disadvantage transmission in terms of channel occupancy since they take longer time than the higher rates to send the same amount of information. This performance anomaly has been presented in [28], it is mentioned that slow host may considerably limit the throughput of other hosts roughly to the level of lower rate. Another constraint in multicast transmission is the lack of acknowledgment and retransmission due to huge amount of traffic overhead these packets will generate. This is severe when transmission mode is multicast because the number of acknowledgment/retransmission will be multiplied by the number of recipients in the multicast group, which could cause collision due to feedback implosion.

The lack of feedback results in two main drawbacks, firstly it is more difficult to know the current situation of the mobile node without feedback mechanism. Therefore, many of the schemes insist the use of feedback mechanism; for example, they make use of RTS/CTS (Request/Clear To Send) frames or channel probing mechanism. Secondly no feedback means no recovery from the loss or error; this makes multicast transmission unreliable. Some researchers have proposed reliable multicast protocols (e.g. [96] or [97]) to deal with unreliability issue. However, this chapter does not focus on reliability problem since it can be assumed that for real-time traffics like UDP-based streaming, reliability is not a crucial issue. It is preferable to lose a few packets than waiting for retransmission, which delays packet delivery. Hence, similar to the previous chapter, the focus here will be on the performance of the network with respect to user satisfaction and network utilization as they are the main objectives of this dissertation.

5.3 Related Works

This section begins with some backgrounds concerning rate adaptation capability in wireless IEEE 802.11 networks. Then, it continues with rate adaptation mechanisms designed for in IEEE 802.11 for unicast and multicast transmission respectively.

5.3.1 Rate adaptation capability

The rise of wireless communications has pushed research and development in this area to grow very quickly. IEEE 802.11 [80] based wireless communications have been widely deployed. Commercial products and numerous access networks are available. Moreover, the standard has provided many specifications for the deployment of wireless networks; one of which is the multi-rate transmission capability provided by 802.11 physical layers. For example: 1, 2, 5.5, 11 Mbps data rates are available in IEEE 802.11b [94]; or 6, 9, 12, 18, 24, 36, 48, 54 Mbps are also available in IEEE 802.11g [98]. These different data rates come from different modulation techniques and channel encoding schemes; for example, in IEEE 802.11b, DBPSK (Differential Binary Phase Shift Keying), DQPSK (Differential Quadrature Phase Shift Keying), CCK (Complementary Code Keying) 5.5 and CCK 11 correspond to data rate of 1, 2, 5.5, 11 Mbps respectively. In wireless environment, different factors such as path loss, fading, or interference in the channel have direct impact on the variation of the received signal to noise ratio (SNR), which results in variation in Bit Error Rate (BER). The lower the SNR the more difficult it is for the modulation scheme to decode the received signal, resulting in higher BER; hence the need of rate adaptation.

5.3.2 Rate adaptation mechanisms in wireless unicast

The first and widely used rate adaptation protocol in commercial products is *Auto Rate Fallback (ARF)* [99]. In ARF, when SNR decreases, an access point tries to recover by decreasing the transmission rate. In fact, the access point switches to a higher rate when a certain number (ten) of packets has been successfully received; it switches back to the lower rate when a failure occurs right after rate increase. If a failure occurs when the number of consecutive successful transmissions is less than ten, the access point switches to a lower rate only after two consecutive failures. Regardless of its wide implementation in commercial products, the protocol has a drawback resulting from the static-threshold approach, which does not adapt well to varying condition in wireless networks.

To solve disadvantages from static-threshold approach, the authors of [100] have proposed *Adaptive ARF (AARF)*. The authors also use threshold-based mechanism as in ARF but instead of setting it to a fixed number, the threshold follows binary exponential backoff continuously at runtime to better reflect to the channel conditions.

This means they multiply by two the number of consecutive successful transmission required to switch to a higher rate. The mechanism increases the period between successive failed attempts to use a higher rate results in fewer failures and retransmissions, thus the overall throughput is improved. Despite that AARF is an efficient mechanism; it cannot be used in multicast transmission since the implementation of this protocol requires acknowledgment and retransmission, which are disabled in multicast.

Another popular protocol is *Receiver-Based Auto Rate (RBAR)* [101] that has the goal of performance optimization in wireless networks using also rate-adaptation mechanism at MAC layer. In RBAR protocol, RTS/CTS mechanism is enabled in order to get/send feedback from receiver. In fact, RTS is sent out before each transmission by the sender and it is received by the receiver who computes the SNR of the frame. After consulting a table mapping of SNR and rate, the receiver sends back the transmission rate for the sender to use in the next transmissions in CTS. RTS and CTS headers have been modified for the purposes. This mechanism is based on SNR (computed with a priory channel model), which is a physical parameter that does not always correlate well with human perception. Moreover, RTS/CTS mechanism is disabled in multicast transmission.

5.3.3 Rate adaptation in wireless multicast

Based on similar idea of using RTS/CTS in RBAR, the authors of [24] have proposed *Rate Adaptive Multicast protocol (RAM)* for channel estimation and rate selection. In this protocol, multicast receivers make use of RTS to measure channel condition and send back transmission rate for sender to use in CTS. In case that a member does not receive the data frame correctly, it will send a NACK (Not Acknowledge). For enhancing the throughput, the authors added a frame sequence field to RTS. This field is used by the member to check whether multicast data frame is a new frame or retransmission. If a frame is a retransmission of a previously successfully received frame, a member will not participate in this multicast transmission. This reduces the number of retransmission. It can be noticed that the protocol makes use of RTS/CTS, NACK and retransmission, which are disabled in multicast. In addition, there are many modifications to existing frames.

To overcome feedback implosion problem, the authors of [26] proposed *Leader-based Rate Adaptive Multicasting for Wireless LANs (LM-ARF)* protocol that deploys leader-based feedback approach and adapts data rate according to ARF. One of the receiving stations, which is the leader, is responsible for sending ACK on behalf of the participating multicast stations. If any multicast receiver, which is not the leader, fails to receive a multicast frame, it will send a negative acknowledgment (NAK) to request retransmission. The AP adjusts the contention window size the same way as that of a unicast transmission thus keeping fairness between unicast flows. New frame type called *CTS-to-Self* frame has been added in order to guarantee the channel access and to announce the transmission of a multicast frame. This mechanism covers several aspects such as fairness, reliability, and performance; however, since it uses ARF, it also inherits the static-threshold approach and drawback of ARF as well.

To avoid using RTS/CTS, the authors of [25] proposed *Auto Rate Selection for Multicast (ARSM)* protocol that uses multicast channel probe operation (MCPO) with multicast probe frame sent out by AP before sending multicast traffic. In this protocol, the user having the lowest SNR will be the one in charge of replying to the AP by multicast response. The AP then selects the multicast data rate based on feedback in three different ways: explicit, implicit, and no feedback. For avoiding collision, multicast users select backoff timer according to their SNR value.

Taking into account user perception, the authors of [27] proposed *SNR-based Auto Rate for Multicast (SARM)*. It adapts transmission rate according to SNR of the node experiencing the worst channel condition. SNR references are obtained from a table listing required SNR for PSNR (peak signal to noise ration) to be higher than 30 (representing good quality) for each transmission rate. By changing multicast transmission rate on the basis of SNR values reported by mobile nodes, the wireless channel is used more efficiently. To overcome the lack of feedback mechanism in multicast, the authors propose a channel probing mechanism to inform the access point of the channel quality at mobile nodes. To avoid collision when nodes transmit feedback to the access point, the author also proposed a backoff timer for each mobile node based on the received SNR. This scheme seems to have the closest objective to the proposed scheme (good user-end quality), thus its results will be compared to those of the proposed scheme.

For better comprehension, we summarized the described schemes in Table 5.1.

Table 5.1: Summary of rate adaptation protocols.

	Protocol	**Threshold**	**Metric**	**Feedback**
Unicast	ARF	static	tx failure	ACK
	AARF	dynamic	tx failure	ACK
	RBAR	static	SNR	RTS/CTS
Multicast	RAM	static	SNR	RTS/CTS, NACK
	ARSM	static	SNR	Channel probing
	LM-ARF	static	tx failure	Leader-based
	SARM	static	SNR, PSNR	Channel probing

5.4 The Proposed Schemes

This section begins with describing how access point gets QoE score in real time. Then it continues with strategy, setup, and results of static and dynamic approaches respectively.

5.4.1 Getting real-time QoE

For a better comprehension of statistics selection for PSQA, some notions of video compression are recalled here. There are three types of frames used in video compression 5.1: I-frames, P-frames, and B-frames. An I-frame is an 'Intra-coded picture', in effect a fully-specified picture, like a conventional static image file. The other two are P-frames (forward-predicted) and B-frames (bi-directionally predicted) holding only part of the image information, so they need less space to store than an I-frame, and thus improve video compression rates.

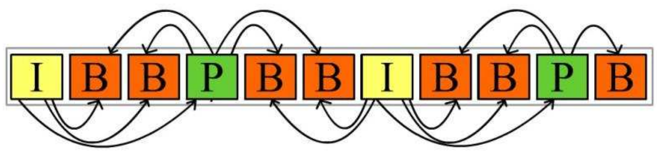


Figure 5.1: Video compression: each P frame is produced from I frame, each B frame is produced from I and P frames.

PSQA has been trained and validated using statistics of application frame level (I/P/B) to map with users' perception. In other words, the following parameters are used: loss rate of the I frame, loss rate of the P frame, loss rate of the B frame, and mean loss burst size (MLBS) of the I frame. The last parameter is used to capture the way losses are distributed in the flow as it affects dramatically the perception of video [86]; this is collected for I frame, which is the most important frame type.

For communications between access point and mobile nodes, the scheme uses IEEE 802.11k standard [102], which specifies many measurement requests and reports that are useful for the proposed schemes. It can be noticed that with IEEE 802.11k measurements, the control traffic is less significant in terms of overhead as it is sent much less frequently than other packet-level schemes. For example, control traffic can be sent every second in this scheme comparing to every single packet in the other packet-level schemes.

An access point in the proposed schemes initiates requests for the actual QoE to users at different timestamps at the beginning of monitoring period in the order of session joining. This is to avoid collision explosion of reports sending back from all users. With PSQA running on every station, users compute their QoE and return it to the access point afterwards. Thanks to this information, when condition changes, the access point adapt its transmission rate accordingly.

5.4.2 Static Approach

Firstly, a static approach is proposed. It is a novel rate adaptation mechanism that adjusts transmission rate according to end-user perception in terms of quality of experience. The idea of the proposed scheme is to use QoE feedback from mobile stations to provision the current condition of the network and then adapt the rate accordingly.

1. Algorithm

QoE indicator is used to switch from one transmission rate to another because it is more relevant to adapt the transmission rate taking into account the quality perceived at the end-user rather than other signal parameters. Also, as explained earlier, the physical modulation plays an important role in such environment and hence adapting modulation would help facing the bad condition.

Assuming PSQA running on every multicast client, Fig. 5.2 illustrates the behavior of an access point in the scheme during multicast session. At the beginning, the access point transmits multicast traffic at its highest rate. The AP monitors its attached clients every monitoring interval (*mi*). Note that this scheme uses time scale in terms of second because this scale is more reasonable than scaling in packet when dealing with human perception. When the timer rings, AP begins by sending requests to multicast members in order of membership precedence in order to avoid collision of reports sending back from members. When a report is received, AP updates the minimum MOS (*min*) of the group accordingly. Once the last report has been received, it compares *min* with the lower bound (*lb*). This *lb* is computed by adding a margin (*mg*) to a reference score (*rf*), which is an acceptable score for the application. If *min* is less than *lb*, then AP switches immediately to one-step lower rate until minimum rate. If *min* is higher than *lb*, then AP increases the counter (representing the duration that AP has been waiting). If the counter reaches a threshold (*th*), then AP switches to one-step upper rate until maximum rate.

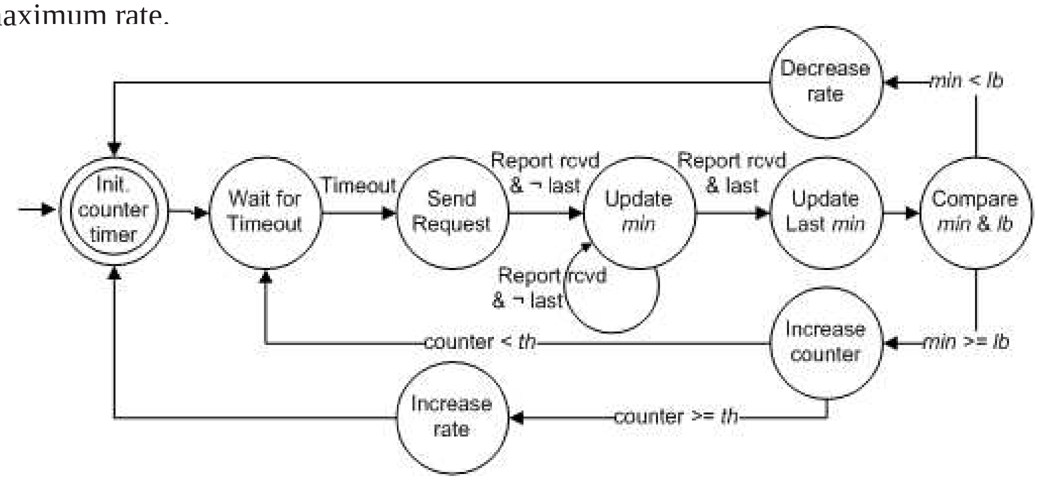


Figure 5.2: Access point behavior during multicast session.

It can be noticed that when condition degrades, the access point in the proposed scheme lowers the transmission rate immediately. This is to adapt instantly to bad condition because it is essential to recover from the bad situation rapidly. When network condition becomes better (i.e. *min* is higher than *lb*) for a certain amount of times, the AP switches to higher rate. This waiting threshold is used to avoid ping-pong effect; before sending at higher rate (higher risk of BER), we should make sure that this condition remains quite stable.

2. Simulation Setup

Firstly, description of the scenario is explained following by those of the implementation. After that, explanation of how to select the value of threshold is given.

- *Scenario*

  Fig. 5.3 illustrates the topology; there is a video server on the Internet with three multicast nodes connected to it via an access point. The encoding rate of the video used in the test is illustrated in Fig.5.4. At the beginning, all nodes locate near by the access point (less than 50m radius). After 10 seconds, station1 (st1) moves away from the access point (150m), and then at the 40^{th} second it begins to move back to its initial position.

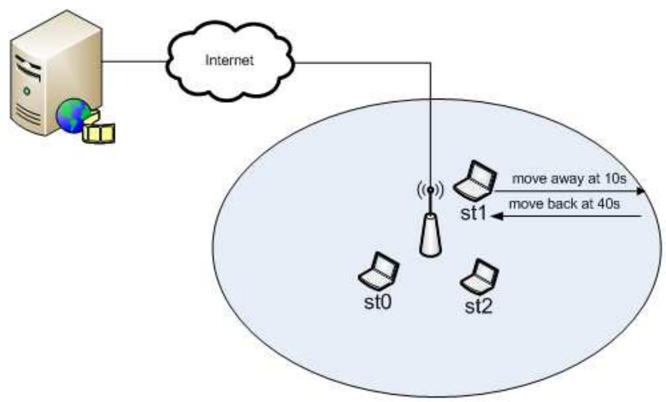


Figure 5.3: Topology of the scenario.

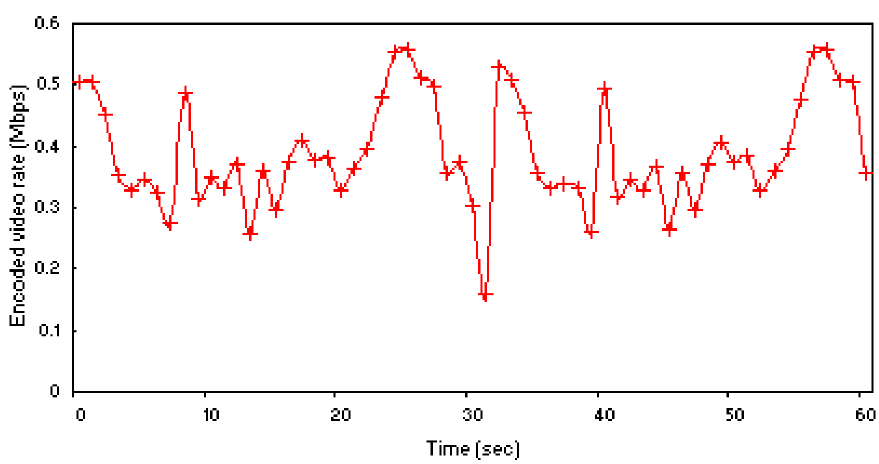


Figure 5.4: Rate variation of encoded video.

- *Implementation of the static scheme*

  Simulation environment is IEEE 802.11 operating in infrastructure mode, as in the previous chapter. The video sequence is an H.264-coded sequence of duration 60 seconds. It is encoded at 384 Kbps and streamed in multicast mode using UDP. The implementation has been done via the network simulator NS-2 [21] version 2.29, patched and integrated with additional modules (`rnn` and `videotrace`) to works with PSQA and to stream real video sequence respectively. In addition, within the modified `mac` module, transmission modulation is adapted automatically according to PSQA score using the proposed algorithm.

- *Threshold Selection*

  Please note that the value of threshold can be set as appropriate to use case. For this example, different values are simulated beforehand to get the best value. Knowing that *mi* is set to 1 second, eight different values for threshold have been tested, ranging from 1 to 8. Fig. 5.5 and Fig. 5.6 illustrate the user experience and the goodput obtained with different values of threshold. Please note that the curves in Fig. 5.5 are normalized, this means that the results are divided by maximum value which is MOS=5. Values in both graphs have been shifted by x which is equal to *i-1* where i is the value of threshold.

  Since the curves have similar trends, which are difficult to interpret; Table 5.2(a) presents summaries of average QoE and goodput values (of all connections) for each threshold. It can be observed that surprisingly the goodput variation is not much affected if the whole connection duration is considered. Therefore, we try to focus on the duration during which the node is in movement (during 20s to 40s); Table 5.2(b) presents these results. With all the arguments observed from the experiments, the value 5 is chosen for *th* because it is a compromised value that gives reasonable reactivity while giving high MOS and goodput. Therefore, the simulations will be conducted with configurations in Table 5.3.

Table 5.2: QoE and goodput of different thresholds.

(a) for the whole connection

Threshold	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
MOS	3.85	4.38	4.40	4.48	**4.48**	4.51	4.59	4.61
Goodput	1.05	1.07	1.04	1.05	**1.06**	1.07	1.07	1.05

(b) during mobility

Threshold	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
MOS	2.98	3.79	3.8	4.13	**4.32**	4.08	4.32	4.51
Goodput	0.98	1.05	0.97	0.98	**1.05**	1.05	1.05	1.06

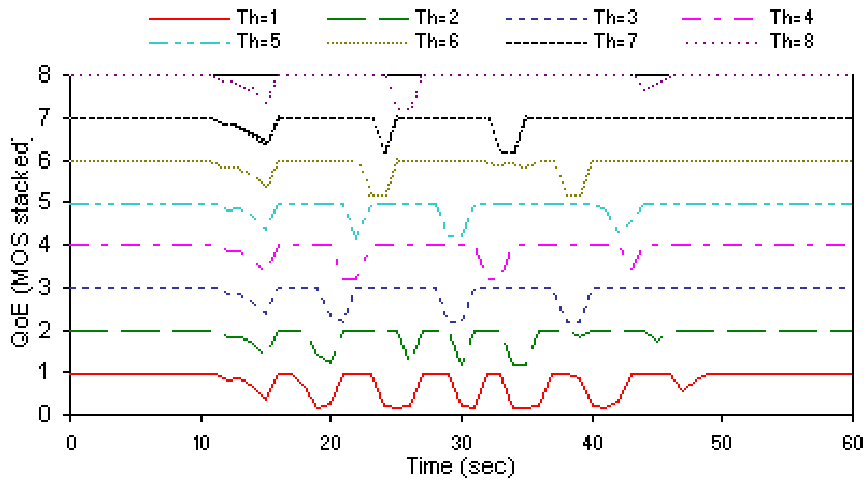


Figure 5.5: Quality of experience for different values of threshold.

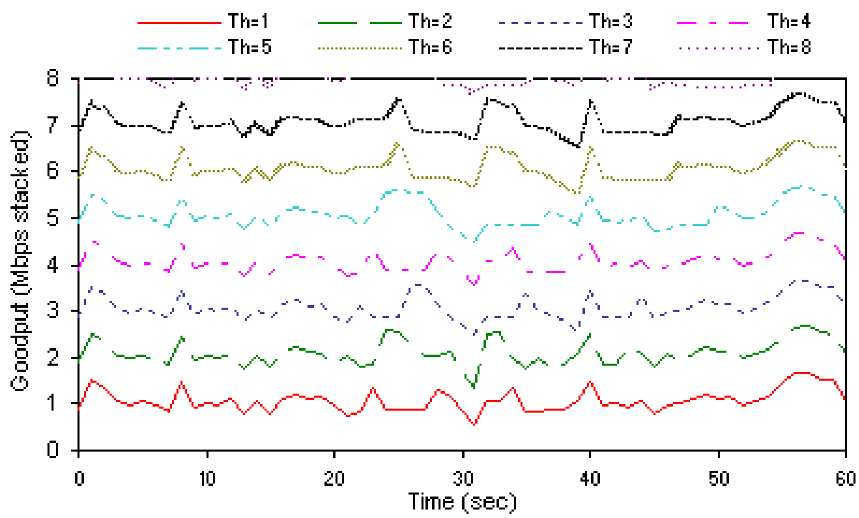


Figure 5.6: Goodput for different values of threshold.

Table 5.3: Configuration of parameters.

Parameter	**Description**	**Value**
mi	monitoring interval	1
th	threshold	5
rt	reference score	3
mg	margin	1
lb	lower bound	4

3. Results

The results are illustrated with two metrics: the goodput (for network utilization) and QoE (for user perception). The proposed scheme is compared to the default multicast (1Mpbs), maximum throughput (11Mbps), and SARM-like mechanism. The comparison of performance with SARM important is because the objective of both schemes is similar. They both want to guarantee quality of service at the receiver; SARM makes use of SNR and PSNR, the proposed scheme makes use of QoE.

- *Goodput*

  Fig.5.7 illustrates the average goodput of all stations obtained from each scheme. Then, details of how individual station behaves in terms of goodput are shown in Fig. 5.8 and Fig. 5.9 for a fixed station (st0) located near by the AP and for a mobile station (moving away from and back to the AP) respectively.

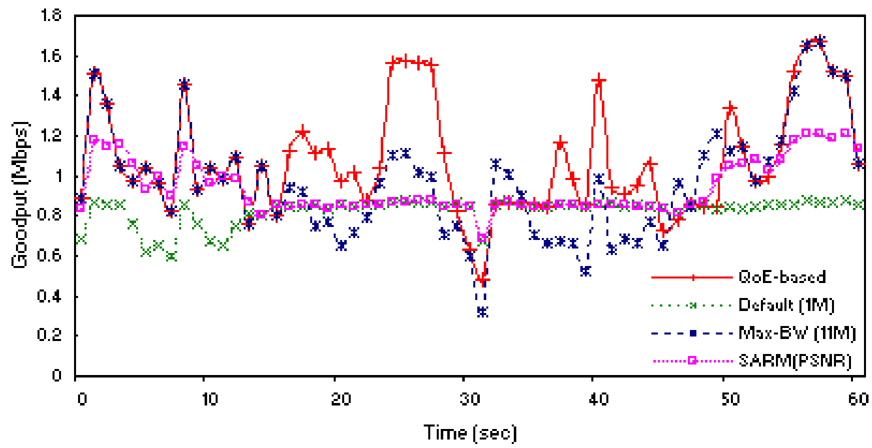


Figure 5.7: Average goodput of all stations.

Observation from Fig.5.7:

- It can be seen that the proposed scheme provides the highest average goodput. More importantly, when the node moves (during 10s to 40s), the average goodput is much higher than all others. This is because the scheme has adapted directly to user perception resulted from several parameters.
- When transmit at default rate (1 Mbps), throughput is the lowest in general (graph before 10s and after 40s). This proves the problem of bandwidth wasting in multicast.
- Using maximum rate gives high goodput at the beginning and at the end; however, when the distance increases (with mobility); channel condition degrades and this strategy performs badly.

– SNR-based performs better than default multicast rate, which is conformed to what have been mentioned in [27]. However, SNR in the scenario is quite low because of mobility and this makes the scheme change to the lowest rate as we can observe in the graph; when the mobile station begins to move, the scheme behaves the same way as in default-1M.

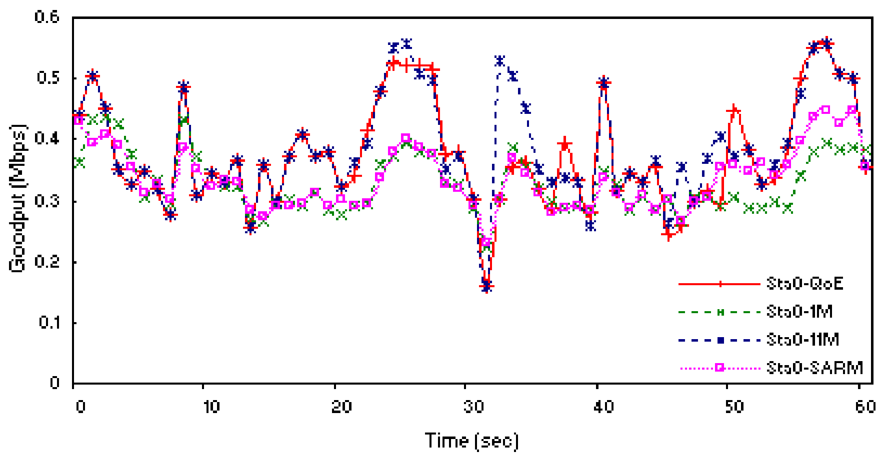


Figure 5.8: Goodput of a fixed station.

We can see from Fig.5.8 that for a fixed station located nearby the AP, its goodput does not change much among different schemes. The variation is due more to the encoding rate (shown in Fig.5.4) than the channel condition. However, we can observe that using 11M for transmission gives a little higher advantage in terms of goodput. This is because when the station is closed to the AP, it can profit efficiently from short distance and high transmission rate.

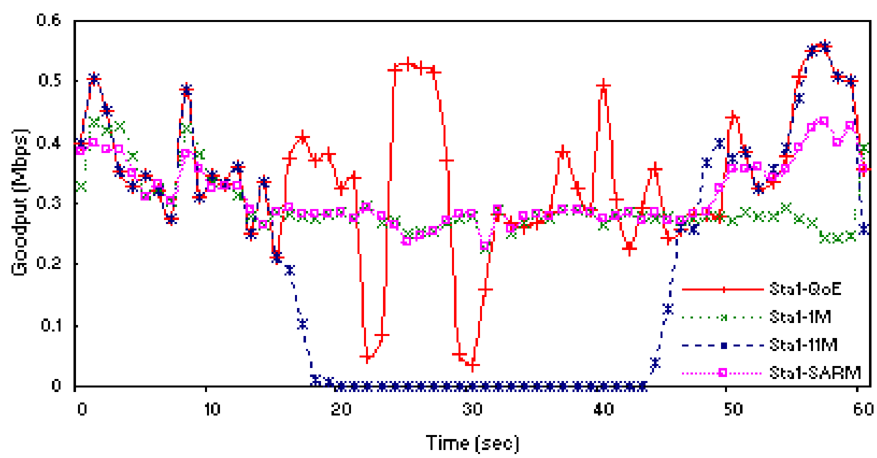


Figure 5.9: Goodput of a moving station.

On the contrary, for a moving station in Fig.5.9 its goodput varies often during station's movement. We observe few drops in the proposed scheme due to the time used to switch to lower rate. We also observe that using high transmission rate (11M) giving very bad results; this is due to the high BER the station suffered when moving away from and back to the AP.

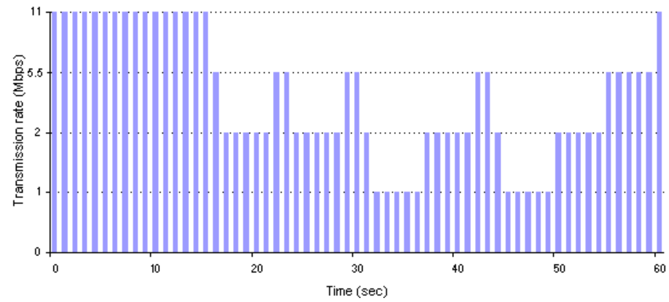


Figure 5.10: Rate adaptation of the proposed scheme during the test scenario.

Note here that this chapter illustrates only the goodput performance of multicast traffic. It can be noticed that if background traffic is also considered, its goodput will be increased when the rate increases and network operator gains more goodput as much as access point transmits at higher rates. This can be explained by the fact that sending at faster rate allows more times for other traffics. The rate variation of the proposed scheme is presented in Fig.5.10.

- *Quality of Experience*

  For QoE performance, two graphs concerning minimum QoE in time and average QoE of all stations are illustrated.

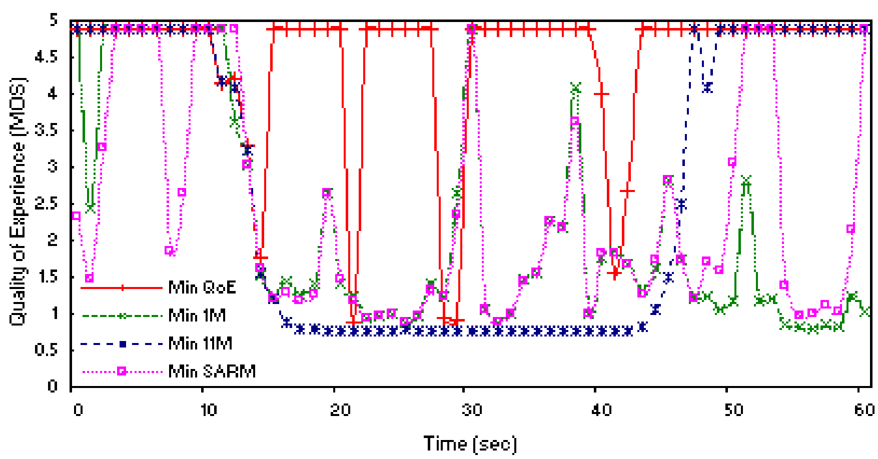


Figure 5.11: Minimum QoE during multicast for each scheme.

Fig. 5.11 illustrates the scores obtained by a member encountered the worst channel condition. It can be seen that the proposed scheme outperforms the others. During moving period, we can see that all schemes experience quite bad performance. The worst scheme is maximum-11M because the rate is too high, and then follows by SNR-based and Default-1M respectively. Despite that the proposed scheme performs the best, we also observe some drops caused by the time taken to adapt to the bad channel condition.

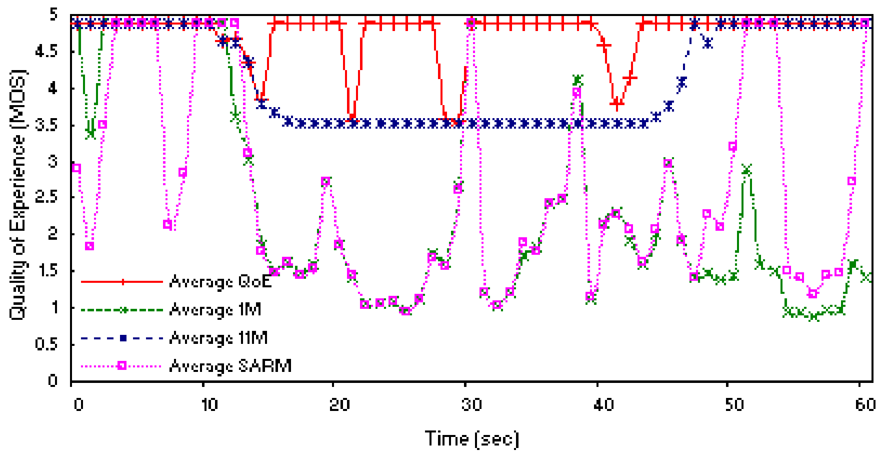


Figure 5.12: Average QoE of all stations for each scheme.

Fig. 5.12 illustrates the overall performance of the network. Since the scheme uses QoE as indicator in the proposed scheme, it gets a great performance in terms of QoE (the average QoE is at least 3.5). However, there are a few drops in the graph due to the time the proposed scheme uses to adapt to the new condition. We also observed that the main problem of SARM-like mechanism may be caused by PSNR definition that does not have a direct relationship with QoE.

5.4.3 Dynamic Approach

We have noticed from previous works that all proposed schemes use a static-threshold rate adaptation for multicast, and none of them has considered dynamic threshold adaptation; the parameters are number of transmission failure or SNR as shown in Table 5.1. The problematic issue associated with static approach is the adaptability to the network condition fluctuation, which is common in wireless environment. Another point to notice is that all the schemes handle rate adaptation according to statistics from packet or frame levels, only SARM uses the concept of PSNR to deal with user perception. Unfortunately, technical parameters do not reveal quality of experience as perceived by the user and it is still questionable whether PSNR has relationship with QoE [103].

In order to overcome different limitations and to adapt to environment and user perception dynamically, *QoE-based Dynamic Rate Adaptive Multicast (Q-DRAM)* is proposed. It is a novel mechanism that dynamically adjusts transmission rate according to end-user perception by mean of quality of experience. It is similar to the static approach because it also uses QoE as metric; however, this mechanism adds dynamic adaptability to help adjusting to network condition dynamically.

1. Adaptation Strategy

The most important decision to make in rate adaptation is mainly on how long to wait (backoff) before changing rate.

- For switching down, the decision is quite simple because we do not want to stay in bad situation so the access point switches rapidly to a lower rate, as seen before. From the literature, there are two causes for switching down. The first one is failure due to varying network condition; this is naturally happened when network condition changes due to mobility, interference, etc. In this case, the sender should wait for two consecutive failures before switching down in order to avoid changing rate up and down all the time (ping-pong effect). The second cause is due to the action that the sender just took to switch to a higher rate; in this case of failure here, the sender switches immediately to the previous rate because the new rate appears to be too high. The proposed scheme switches down immediately after both cases. Note that failure in this scheme occurs when QoE is less than a desired threshold (more details in next subsection). The ping-pong effect will not affect the proposed scheme since it uses time scale in unit of second, which is long enough to avoid it.

- For switching up, the scheme uses dynamic-threshold strategy called binary *exponential backoff (BEB)* similar to AARF. This strategy allows us to adapt to varying network condition. With BEB, access point increases the backoff exponentially when failure occurs or repeats after the successful attempt of rate increase. It means that if the QoE is less than desired (*fail*) right after switching up (*just_up*); the access point switches down immediately and before trying to switching up again it waits longer by setting the backoff timer to be twice of the previous value. For the other case of failure (varying condition), the scheme does not update backoff stage. Fig.5.13 illustrates how BEB works in the proposed scheme. At the beginning, the backoff timer is set to minimum value (*thMIN*). During multicast session, it will be reset to *thMIN* again after a successful attempt of rate increase. The backoff timer cannot exceed *thMAX*. Therefore, the backoff timers corresponding to each stage in Q-DRAM are {0:1; 1:2, 2:4}.

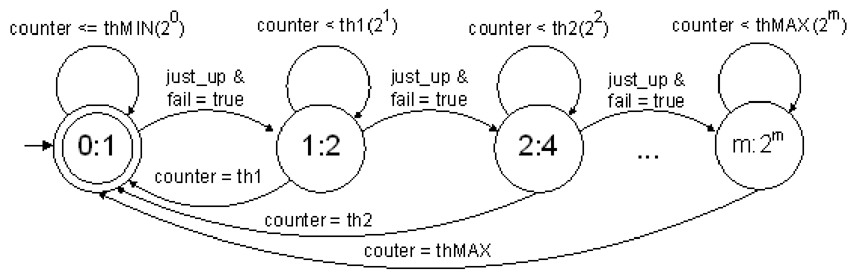


Figure 5.13: Binary exponential backoff in Q-DRAM.

2. Access Point's Algorithm

Fig. 5.14 illustrates the behavior of an access point during multicast session. PSQA is assumed to be running on every multicast node. At the beginning, the access point transmits multicast traffic at its highest rate. The AP monitors its ongoing clients every monitoring interval (*mi*) in unit of second. When the timer rings, AP begins by sending requests to multicast members in the order of membership precedence (to avoid feedback implosion). When a report is received, AP updates the minimum MOS (*min*) of the group accordingly. Once the last report has been received, it compares *min* with the desired QoE called as lower bound (*lb*), computed as in the static scheme. If *min* is less than *lb*, then AP switches immediately to one-step lower rate until minimum rate; in case of unsatisfied QoE just after rate increase, the backoff stage is updated. If *min* is higher than *lb*, then AP increases the counter (representing the duration that AP has been waiting); if the counter reaches a threshold (th_i) where i is backoff stage, then AP switches to one-step upper rate and the backoff stage is reset.

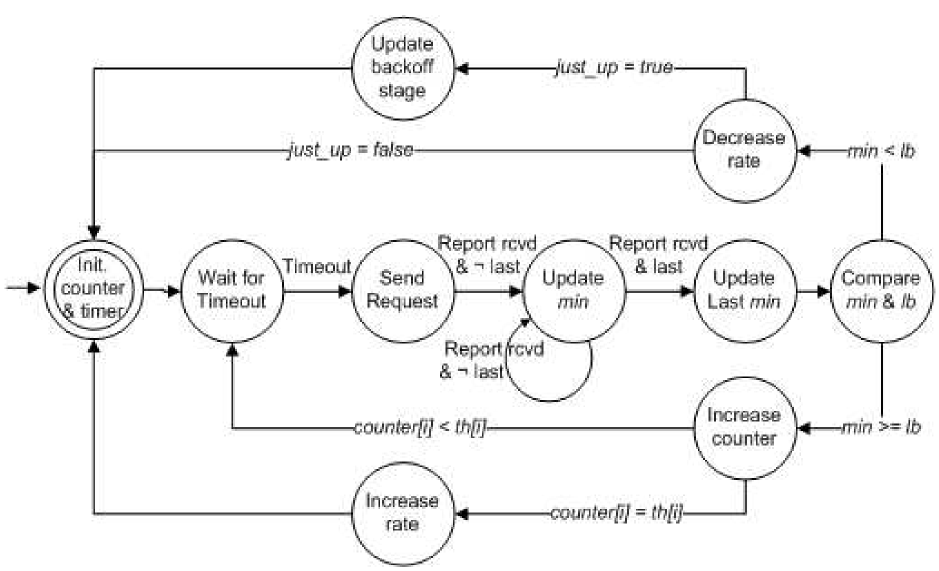


Figure 5.14: Access point behavior during multicast session.

3. Simulation Setup

For the simulation, setup is the same as in static approach. The scenario is briefly recalled here, all nodes are located near by the access point (less than 50m radius) at the beginning, after 10 seconds, station1 (st1) moves away from the access point (150m), and then at the 40^{th} second it begins to move back to its initial position. After extensive simulations, the scheme uses *thMIN*=1s, and *thMAX*=4s. Hence, backoff timers corresponding to each backoff stage are {0:1; 1:2, 2:4}. The value of *thMAX* is set to 4s in order to react rapidly to condition change. Implementation has been modified according to the dynamic algorithm.

4. Results

Results are demonstrated in terms of goodput and QoE. The proposed scheme is compared to 1Mbps, 11Mbps, and SARM-like mechanism as before.

- *Goodput*

  First, Fig.5.15 illustrates the average goodput of all stations obtained from each scheme. Then, two more graphs are presented: a fixed station (st0) located near by the AP in Fig. 5.17 and a mobile station (moving away from and back to the AP) in Fig. 5.18. Note that the goodput is normalized according to the encoding rate of the video thus the obtained results scaled in the interval [0:1].

  It can be seen from Fig.5.15 that the proposed scheme provides the highest average goodput. More importantly, its average goodput is significantly higher than all others schemes during node movement (10s to 50s); however, we also observed the fluctuation generated by attempts of rate increasing during this period. We observed similar behaviors for 1M and 11M schemes. For SARM, it

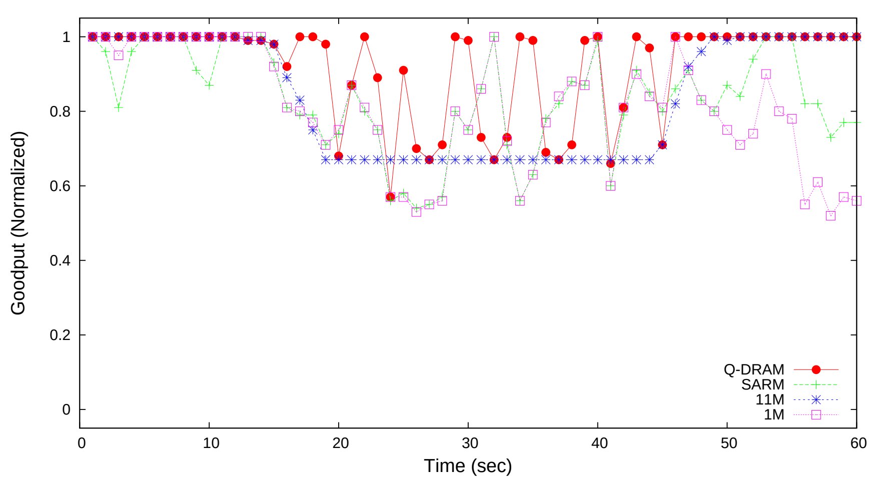


Figure 5.15: Average goodput of all stations.

performs slightly better than basic rate in general; even so, there is no improvement during period of mobility.

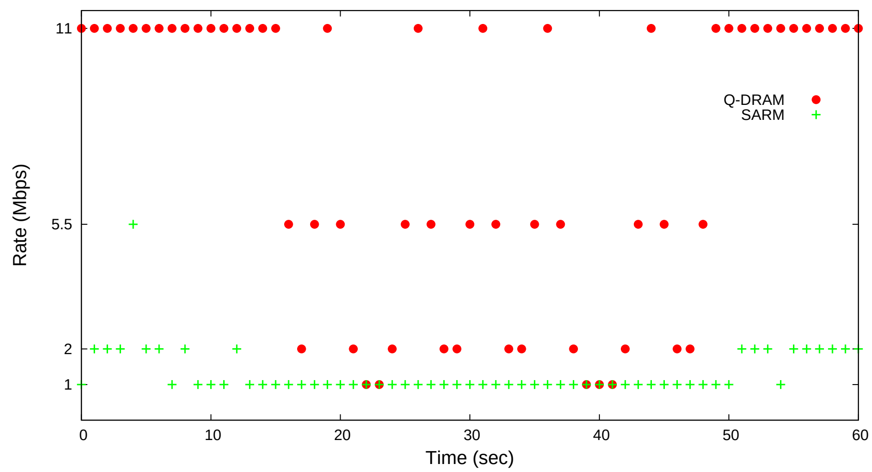


Figure 5.16: Selected rates of Q-DRAM and SARM during the test scenario.

The results obtained in Fig.5.15 are confirmed in Fig. 5.16, in which we present selected rates of Q-DRAM and SARM during the simulation. We can notice that Q-DRAM uses high transmission rates when possible; which results in better goodput comparing to SARM. During mobility (low SNR), the proposed scheme attempts to increase rate as soon as it detects good channel condition, hence the consequence in several rate switching. But again, Q-DRAM still outperforms the other schemes during this mobility period.

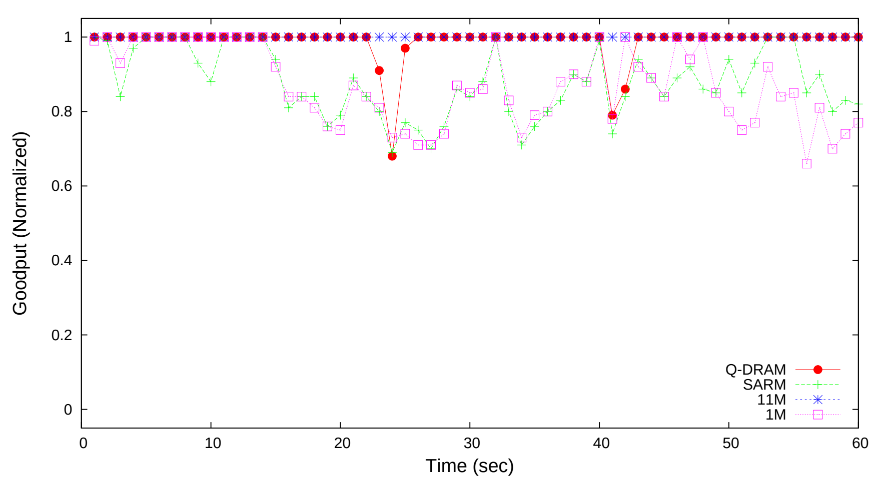


Figure 5.17: Goodput of a fixed station close to AP (st0).

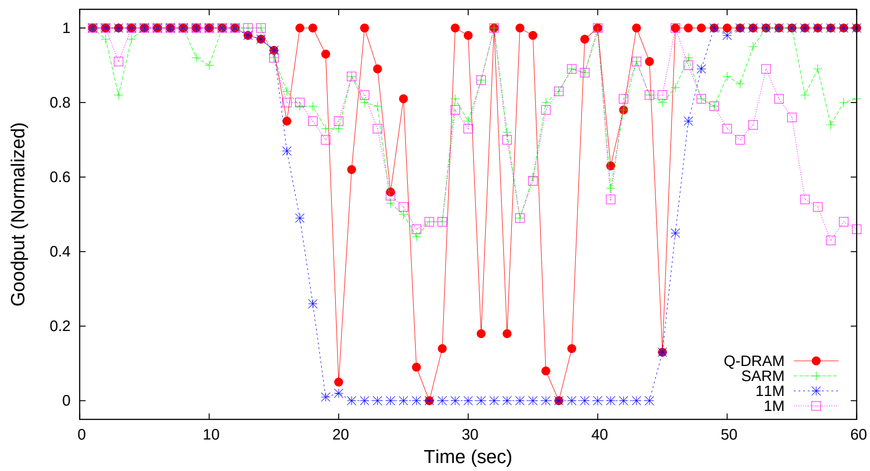


Figure 5.18: Goodput of a moving station (st1).

For a fixed station situated near by the AP (Fig.5.17), the goodput is excellent when using 11Mbps and Q-DRAM. This is because the station closed to the AP can profit efficiently from short distance (high SNR), which allows us to use relatively high transmission rate. On the other hand, for a moving station in Fig.5.18, the goodput varies often during station's movement. We observe some drops in Q-DRAM due to failed attempts to increase rate. We also observe that 11Mbps gives the worst performance as the rate is too high; this generates high BER and high Frame Error Rate (FER) seen in Fig.5.19. However, we noticed that even when using the lowest rate as in SARM and 1Mbps, the goodput also stays in bad situation.

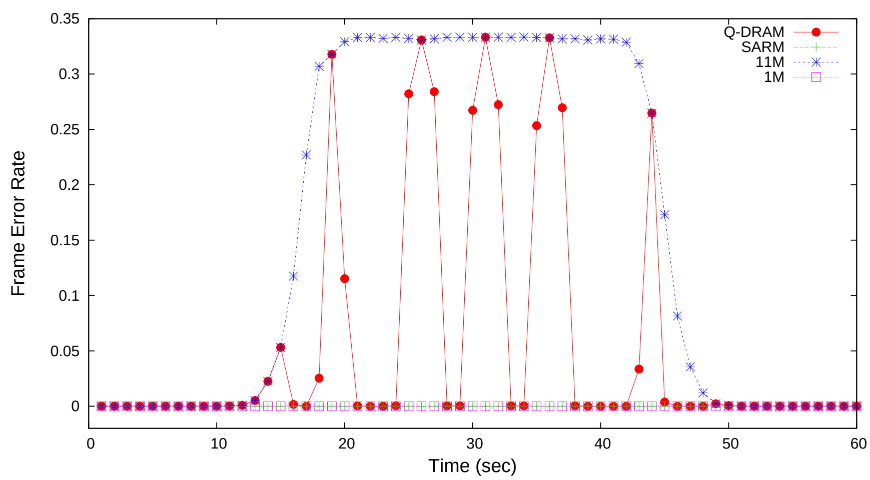


Figure 5.19: Average FER of all stations for each scheme.

- *Quality of Experience*

  Fig.5.20 presents the scores obtained by a member encountered the worst channel condition (lowest MOS), of which the scheme took as reference for adjusting rate. It can be seen that Q-DRAM performs the best regardless of some drops resulting from failed attempts during mobility. In order to gain more throughputs the scheme prefers to try to switch often. As seen before, 11Mbps performs the worst during mobility because of high rate. SARM and 1Mbps give similar performance as SARM has adapted to use 1Mbps during node movement; even so, this rate is not fast enough to transmit all encoded data.

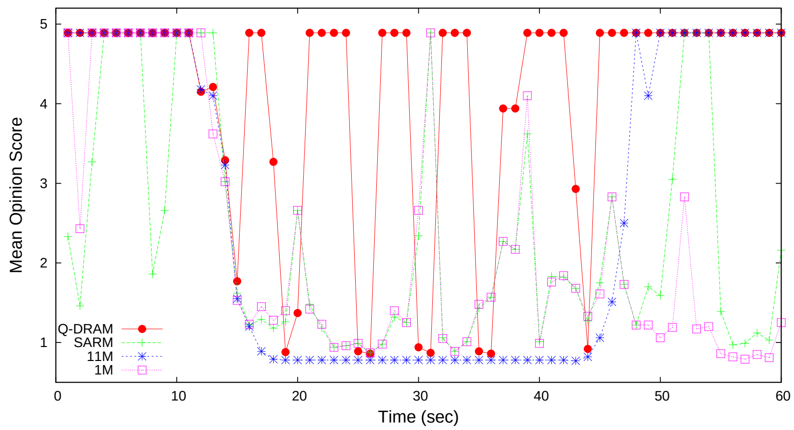


Figure 5.20: Minimum QoE obtained from st1 for each scheme.

Finally, Fig. 5.21 illustrates the overall performance of the network regarding user satisfaction by mean of average MOS of all stations. It can be noticed that Q-DRAM obtained a great performance in QoE (the average MOS is at least 3.5 along the session). However, there are a few drops in the graph due to the failed attempts of rate increase.

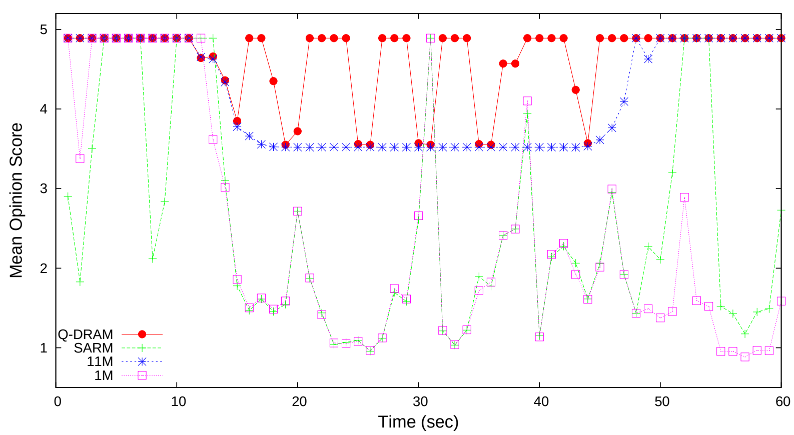


Figure 5.21: Average QoE of all stations for each scheme.

5.5 Conclusions

This chapter deals with problems of rate adaptation in wireless multicast. The reason why the rate adaptation mechanism is used to treat multicast performance problem is because all losses in multicast mainly resulted from channel error. Thus, adapting transmission rate can help improving network performance.

Studies have been carried out for different approaches and we have seen from the results that using default basic rate for multicast transmission (conservative approach), has drawbacks not only in terms of network utilization but also in quality perception at the users. Deploying the maximum rate (11Mpbs) gives great performance if network condition is good, however when the condition degrades this high rate leads to poor performance. Many schemes, including SARM, make use of PSNR as metric for changing rate but PSNR does not always imply accurately user experience, which is essential in real-time multimedia applications.

To obtain good network utilization while maintaining user satisfaction, two novel rate adaptation mechanisms have been proposed (one with static strategy and another with dynamic strategy). They are both based on quality experienced by multicast clients for rate adaptation. It can be noticed that for mechanism like rate adaptation, the threshold indicating when to switch transmission rate is the heart of mechanism; therefore, the method to select the threshold is very important. In static approach, the fixed threshold-based mechanism with quality of experience as metric has been investigated. Different values for selecting the best threshold have been studied and the obtained results are satisfying in both QoE and goodput.

For a better adaptability to varying network condition, the dynamic approach is also proposed. Deeper investigation has been conducted on threshold selection deploying adaptive strategy (use of binary exponential backoff). The scheme is dynamic, it can thus adjust transmission rate according to varying wireless condition better than a static approach. As a consequence, it also achieves good performances in both network utilization and user perception.

Chapter 6

Packet Scheduling

6.1 Introduction

Today multimedia applications can be supported under various technologies. As we have seen how QoE can work in WLAN, this chapter will explore Cellular networks, another popular network technology with support on high mobility. The focus is on one of them called Universal Mobile Telecommunications System (UMTS). Improved with a new access method (High Speed Downlink Packet Access or HSDPA), it can provide higher bandwidth and enable wider range of services including multimedia applications. In UMTS, different categories of traffic are specified along with their characteristics. Best effort traffic has been specified with low priority because it has fewer constraints. On the other hand, real-time multimedia traffic such as streaming video or VoIP are more sensitive to network condition changes, hence special treatment (e.g. QoS scheduler) is needed in order to achieve user satisfaction.

In order to reach this goal, an efficient *packet scheduler* is necessary. According to the literature, most of scheduling mechanisms mainly take into account signal quality and fairness and do not consider user perception. In this chapter, a novel approach is presented with QoE-aware schedulers that take quality of experience into account when making scheduling decisions. The remaining of the chapter is organized as follows. First, background on UMTS is described in Section 6.2, then Section 6.3 gives related works, meaning existing schedulers in HSDPA. Section 6.4 describes QoE-aware scheduling mechanism and Section 6.5 presents description implementations and scenarios. Section 6.6 presents the obtained results considering various schedulers and parameters. Finally, Section 6.7 provides conclusions and future works.

6.2 Universal Mobile Telecommunications System

Universal Mobile Telecommunications System or UMTS [104] is a third-generation (3G) wireless cellular network that offers higher data rates than older 2G and 2.5G mobile networks. A typical UMTS network is shown in Fig. 6.1. The Figure shows a core network and the UMTS Terrestrial Radio Access Network (UTRAN). The UTRAN consists of Radio Network Controllers (RNC) which control several base stations (BS). A mobile user with her connected User Equipment (UE) to the UTRAN can communicate to other networks like the Internet, through Serving GPRS Support Node (SGSN) and Gateway GPRS Support Node (GGSN) located in the core network.

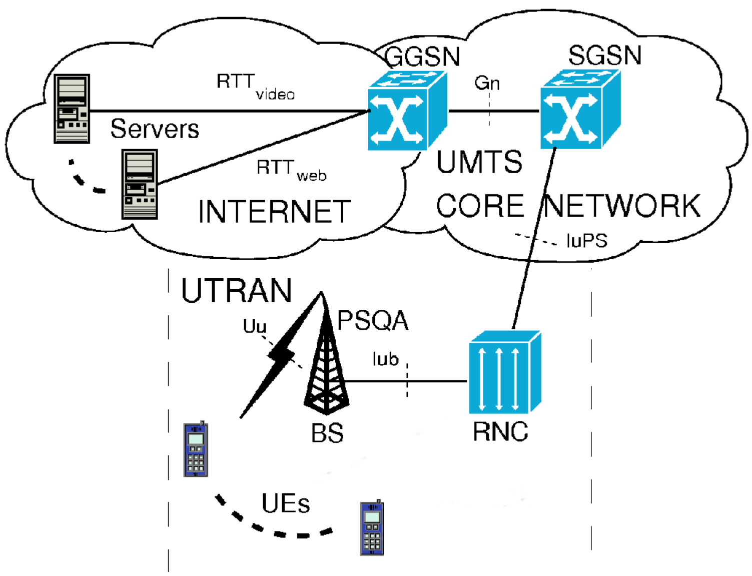


Figure 6.1: Simulation topology.

To fulfill different QoS requests, four different QoS classes (conversational, streaming, interactive, and background) are specified in UMTS along with their fundamental characteristics. Delay sensitivity is used as the main distinguishing factor. Conversational class is the most sensitive while background traffic is delay insensitive. Conversational and streaming are Real-time (RT) whereas interactive and background are non real-time (NRT). The real-time traffic does not tolerate well the delay because of bi-directional communication, for example in voice or video telephony over IP. The last two categories are interactive and background class. They are both best effort traffic, example of interactive traffic is web browsing and background traffic is email. The background class has less priority since the receiver does not expect the data within a strict delay and thus some delays can be tolerated. Streaming technology is becoming increasingly important due to following reasons. First, as storage capacity in a mobile device is much less than in a computer, user cannot store the whole file before playing it. Second, most users do not have fast enough connection to download entire file quickly. With streaming they can start displaying the media before the entire file has been transmitted. Last but not least, there exist different types of streaming, one of which is real-time streaming that is becoming more and more popular today. For streaming, a steady and continuous connectivity is necessary in order to get a good service quality at end-user.

With the Release 5 of UMTS, 3GPP started the work on High Speed Packet Access by specifying the so-called High Speed Downlink Packet Access [104] that supports data rates of the order of 10 Mbps. The increased bandwidth provided by HSDPA enables the deployment of a wide range of services, like voice, data, and multimedia streaming. In particular, video streaming services are becoming popular and will likely be a significant source of revenues for UMTS operators. In HSDPA [104], fast monitoring of the radio channel conditions of all users is performed; at every Transmission Time Interval (TTI) of 2 ms, a UE can send a Channel Quality Indicator (CQI), to the BS, over a control channel. Such feedback makes it possible to adapt the coding rate, modulation scheme, and number of codes employed, so that users having good channel conditions may be provided with high data rates. Therefore, the scheduler chooses every TTI the next user to be served based on the channel conditions of *all* the users, and possibly also their different QoS requirements. Reader can refer to [104] for more background on HSDPA.

With increasing amount of multimedia traffic running on UMTS today, quality definition has also been shifted from quality of service to quality of experience. Within UMTS network, many works have been done regarding QoE measurement but very few on QoE management (e.g. [105, 106]). To the best of my knowledge, this document provides the first investigation of possibility to use QoE as metric for scheduling decision in UMTS.

6.3 Related Works

HSDPA scheduler is the key to resource management in the UTRAN downlink, because it decides which user is to be scheduled at each time slot. Many scheduling methods have been proposed and some of the representative methods are describe here.

- *Round-Robin (RR)* is one of the simplest schedulers. It gives the time slot to the users in a round-robin manner and is fair with respect to system resources (time slots). However, this policy is not optimal in terms of system throughput as it does not take into account users' channel conditions and QoS requirements of application.

- *Maximum Carrier-to-Interference Ratio (CI)* gives the channel to the user having the best channel conditions at each given time slot. If $R_i(t)$ is the instantaneous data rate experienced by user i at time t, then the CI scheduler assigns the slot at time t to a user i^* such that $i^* = \arg\max_i \{R_i(t)\}$. That is, it gives the channel to the user able to achieve the highest instantaneous rate. The CI scheduler provides the highest system throughput but it is very unfair as a user closer to the base station can get all the resources, and the users farther away (bad CQI) will have to face starvation.

- *Proportionally Fair (PF)* [107] assigns the slot at time t to a user i^* such that $i^* = \arg\max_i \{R_i(t)/\lambda_i(t)\}$, with $R_i(t)$ the same as for CI and $\lambda_i(t)$ is the *average throughput* of user i:

$$\lambda_i(t) = (1 - 1/\tau) \cdot \lambda_i(t - \Delta t) + 1/\tau \cdot R_i(t). \tag{6.1}$$

  Here, $\tau > 1$ and Δt is equal to the length of the TTI. The PF scheduler offers a good trade-off between system throughput and fairness, as it both gives the channel to the user having "relatively good" channel conditions and hence provides the so-called "proportional fairness" defined in [108].

- *QoS schedulers* [109, 110, 29, 30] try to satisfy some QoS requirements such as guaranteed throughput, minimum delay, etc. QoS schedulers in general pick a user i^* satisfying

$$i^* = \arg\max_i \{B_i(t)R_i(t)/\lambda_i(t)\}, \tag{6.2}$$

  where $B_i(t)$ represents a "barrier function" [29]. A QoS scheduler called *Normalized Rate Guarantee (NRG)* [30] that in turn is based on Rate Guarantee (RG) scheduler [29] is considered in this chapter for performance comparison. NRG scheduler is given by equation (6.2) where $B_i(t)$ for QoS users $\mathcal{Q}$ and Best Effort users $\mathcal{B}$ is given by:

$$B_i(t) = \begin{cases} \lambda_{min}^{(i)} + \lambda_i(t)\beta \cdot \exp\left(-\beta \cdot \frac{\lambda_i(t) - \lambda_{min}^{(i)}}{\lambda_{min}^{(i)}}\right) & \forall i \in \mathcal{Q}, \\ k_{BE}/n_{BE} & \forall i \in \mathcal{B}. \end{cases} \tag{6.3}$$

  where λ_{min} is the guaranteed rate and n_{BE} is the number of Best Effort users. Moreover, k_{BE} and β are engineering parameters and tuning them involves the tradeoff between NRG adhering strictly to rate guarantees or higher overall throughput for best effort users. Similar to [30], the values of k_{BE} and β are taken to be 1500 and 6.0 respectively when rate units are in kbps.

  NRG provides rate guarantees to QoS users and it improves upon RG such that it apportions losses in a fairer way during congestion irrespective of different rate guarantees and unlike RG it avoids deteriorating QoS when BE load increases. In [30], NRG is evaluated using a QoE estimation module. Possible adaptation strategies that can use the QoE feedback are not investigated. Thus, the current chapter studies the use of QoE feedback for adaptive packet scheduling in HSDPA.

6.4 QoE-aware Scheduler

This section gives explanation of how the QoE-aware schedulers work. It begins with description of real-time QoE assessment, and then scheduling algorithm is described; it explains how the scheduler selects a station to be scheduled.

6.4.1 Real-time QoE monitoring

In order to get quality of experience feedback in real-time, *PSQA* tool [16] is used. Every t milliseconds, PSQA obtains the required parameters from the received video packets, over a play-out window of T_w in the past, and uses them to estimate MOS in real time. PSQA module is placed at BS so that the scheduler can get MOS scores for making scheduling decision. The scheme assumes that BS has knowledge of packet loss statistics either via upper layer sequence numbers or another mechanism that can be implemented in BS itself. In the simulations, t = 24ms and T_w = 5s. To obtain a single PSQA score of the entire video for plotting the graphs, the scheme takes the average of all the MOS scores obtained over time.

6.4.2 Algorithm of the scheduler

Two algorithms are proposed; the first one is called *QoE-CI*. It is based on CI with the objective of maximizing system throughput while taking into account the quality of experience of video-streaming users. The second algorithm is called *QoE-PF*. It is based on PF; the goal of this algorithm is to maximize fairness between users while keeping QoE of video users acceptable. Both schemes use a variable called *threshold (th)* that can be tuned by network operator. This variable represents the quality threshold that is desired by video users. In the test, threshold variations are {3.0, 3.5, 4.0, 4.5} considering that an acceptable value of QoE is 3 ("Fair" quality). In order to behave closely to original schedulers, *QoE strategy* is applied only when MOS_{min} (minimum QoE score at every t interval) is less than the desired threshold *th*.

The main idea of QoE based strategy is to give higher priority to video users, who have higher constraints in terms of quality. For that, a coefficient is assigned to each user. This coefficient, in a way equivalent to $B_i(t)$ in equation (6.2), is to be multiplied to the priority index ($R_i(t)$ for CI and $R_i(t)/\lambda_i(t)$ for PF), used in each traditional scheduling scheme. The schemes differentiate the computation of coefficient between background and video streaming traffic as following. Note that here QoE scores, MOS_{min} and thresholds are normalized to scale [0,1] before the computation.

- For a background user, $coef\_bg$ is the coefficient of all users; MOS_{min} is the minimum MOS value of all video users. The value of coefficient for each user is then:

  $$coef\_bg = 1 - (1 - MOS_{min}), \tag{6.4}$$

  thus, this scheduler will pick a user i^* satisfying:

  $$i^* = \arg\max_i \{coef\_bg \cdot R_i(t)\}. \tag{6.5}$$

  This implies, the lower MOS_{min}, the lower the chance that background user will be scheduled in the next time slot.

- For a video user, $coef\_vdo_i$ is the coefficient of user i and MOS_i represents current quality of experience of this user within current T_w. This avoids ping-pong effect that could occur if we only measure instantaneous score at t. The value of coefficient for each user is then:

  $$coef\_vdo_i = 1 + (1 - MOS_i), \tag{6.6}$$

  thus, this scheduler will pick a user i^* satisfying:

  $$i^* = \arg\max_i \{coef\_vdo \cdot \{R_i(t)/\lambda_i(t)\}\}. \tag{6.7}$$

  The video users are privileged over background users because they are more sensitive to quality degradation. The lower MOS_i, the higher the chance that the video user i will be selected in a given time slot.

It can be noticed that while considering signal quality and average throughput of each user, the coefficient is added to them when minimum QoE score is below the threshold. This will give higher priority for video stations in degrading situation. For background traffics, since the delay constraint is less important, they can wait for next time slots.

6.5 Performance Evaluation

For performance evaluation, the focus will be on three important metrics namely *quality of experience*, *throughput*, and *fairness*; with variations of schedulers, number of users, and distance from the base station. In order to simulate UTRAN, the EURANE extensions [22] to NS-2 is used. EURANE simulates the RLC (Radio Link Control Protocol) and MAC-hs (MAC in HSDPA) in detail. The RLC layer consists of two modes of operation, Unacknowledged Mode (UM) and Acknowledged Mode (AM). There are per-user queues in the RNC and Droptail queuing is used. The MAC layer implements the HSDPA schedulers. All considered scenarios correspond to the network topology shown in Fig. 6.1. There are fixed numbers of video users and background TCP flows corresponding to users downloading large data files.

The video is a well-known H.264-coded reference sequence called “mother and daughter”. The size format is QCIF and the video is repeated 2 times to make its duration equal to 60 seconds. This duration is more than enough to utilize the PSQA feedback by any resource management module; also, the recommended video lengths for subjective testing ($\approx$ 10 s) are much less than that duration. The average bit rate ($\approx$ 384 kbps) of the encoded video is controlled using the quantization parameter of the codec. A GOP (group of pictures) size of 16 frames is used. The trace of the encoded video file is used for NS-2 simulations. During the simulation, a PSQA module is running at BS and it computes the relevant parameters to obtain a MOS estimation in real-time. For a given scenario with a specific set of studied parameters, independent runs are performed for at least 20 times. QoE scheduling is implemented in the `umts` module of the simulator according to the description. Table 6.1 summarizes the default values of simulation parameters that remain unchanged unless specified otherwise.

Table 6.1: Default simulation parameters.

Parameter	**Default Value**
EURANE configuration & channel model	See [22]
BS transmission power	10 W
Multipath fading environment	Ped A, 3.0 Km/h
λ_{min}, τ	400 kbps, 1000
RLC mode, RNC queue size L	UM, 128 IP packets
Round Trip Times:	
RTT_{video}, RTT_{web}	100 ms, 40 ms to 200 ms
$RTT_{Gn}, RTT_{IuPS}, RTT_{Iub}$	20 ms, 1 ms, 30 ms
Simulated time (single run)	60 seconds
QoE measuring interval (t)	24 milliseconds
QoE measuring window (T_w)	5 seconds

6.6 Results

In this section, performance evaluation results are discussed beginning with QoE scores and throughput obtained in each scheme when varying *threshold*, *distance*, and *number of background traffic* respectively. After that, the fairness issue is discussed at the end of this section.

6.6.1 Threshold variation

As described in the previous section, threshold *th* can be tuned by network operator as the desired QoE value. This section presents an evaluation with different values of this threshold (th=3.0, th=3.5, th=4.0, th=4.5). There are 4 video nodes and 8 background nodes (FTP) in the topology and their maximum distance from the BS is 300 meters. This distance is chosen because the quality becomes very bad beyond 300 m with the encoding rate of 384 kbps (cf. subsection B). Fig. 6.2 summarizes minimum QoE scores, which is the value such that 95% of the video users across all simulation runs get a score higher than this value. For reference, the figure also presents minimum QoE scores obtained by other schedulers.

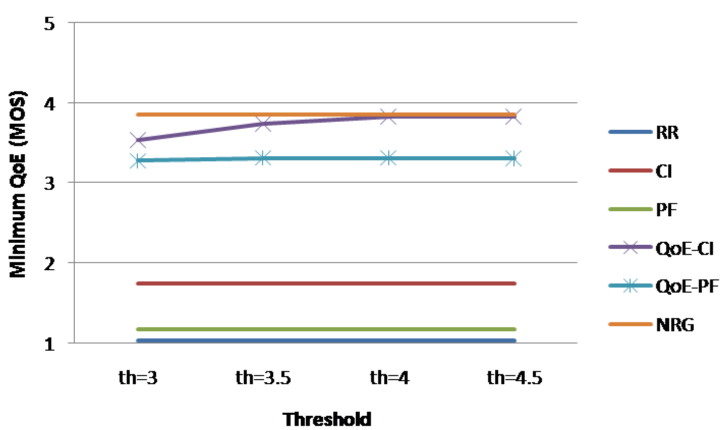


Figure 6.2: Minimum QoE for different threshold values.

It can be seen that QoE-aware schedulers provide great improvement comparing to traditional schemes (RR, CI, and PF) as the minimum scores are always higher than 3 whereas the others remain between poor and bad quality. However, NRG performs slightly better than the proposed schemes but this performance comes with a cost in terms of throughput that will be discussed later. We also observe that the QoE-CI score increases when the threshold increases but the minimum scores do not improve much when *th* is higher than 3.5, therefore, *th* is set to 3.5 for the following investigations. As for QoE-PF, the minimum scores obtained are quite stable while varying threshold value. After investigation, this can be explained by the fact that since the scheme takes minimum scores, the variation of threshold does not have much impact on quality as network condition is very poor. Moreover, no improvement can be done in this type of situation.

Fig. 6.3 represents QoE scores during simulation time. Here, different schedulers are used, the threshold *th* is set to 3.5 for QoE-aware schedulers. We can observe that NRG, QoE-CI, QoE-PF performs well by giving scores always higher than 3 where as the traditional schedulers have bad performance. Note that background traffic starts after 10 seconds.

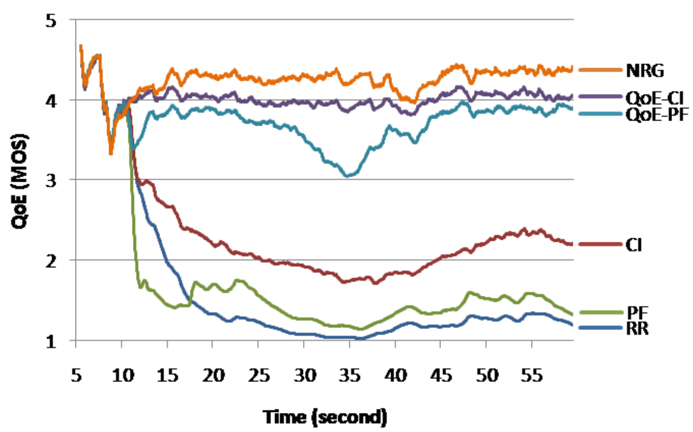


Figure 6.3: QoE scores of different schemes when th=3.5 in QoE schedulers.

6.6.2 Distance variation

In this scenario, simulations are run with variation of the distance (distance of each node from the base station) while using 4 video nodes and 8 background nodes. Also, the QoE threshold (th) is 3.5 and a set of maximum distances is (*100m*, *200m*, *300m*, *400m*, and *500m*). For each run, the distance for each node to the BS is randomly chosen with the maximum distance value configured from this set. Fig. 6.4 presents average scores obtained with all schedulers. We can notice that QoE-aware schedulers perform very well by ensuring average QoE higher than 3 for all distances whereas scores in RR, CI, and PF are much less. However, QoE scores obtained by QoE-aware schedulers are slightly lower than NRG, yet the difference of less than 0.5 is hardly detectable by user. As mentioned before, the performance of NRG comes with a cost which is the throughput; we can see in Fig.6.5 that NRG gives the smallest throughput among all schemes. With advantage in terms of QoE but drawback in throughput, one should consider this tradeoff when choosing a scheduler. Also, please note that the distance beyond 300 m results in bad quality and hence the reason why distance of 300 m is used for the rest of simulations.

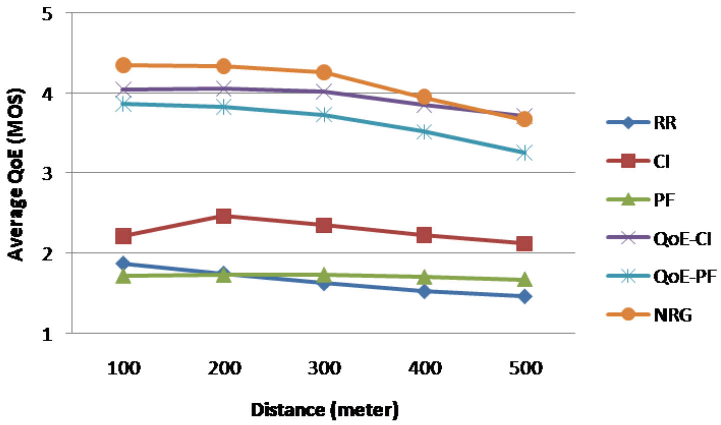


Figure 6.4: Average QoE for different distances.

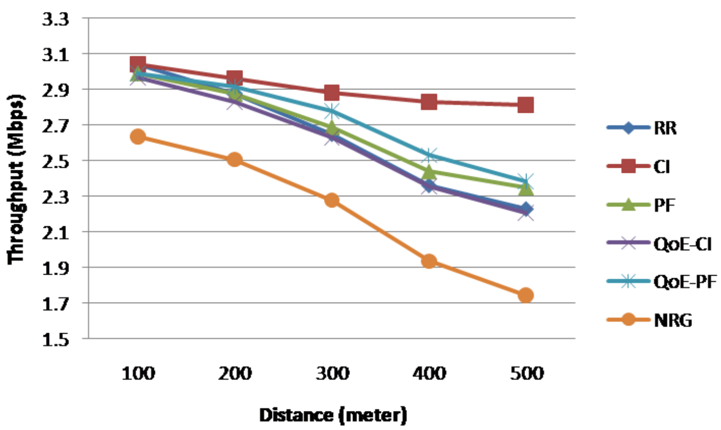


Figure 6.5: Global throughput for different distances.

6.6.3 Background traffic variation

More investigations are done using scenarios with various number of background nodes (4,8,12,16,20) while keeping 4 video nodes, th=3.5 and maximum distance=300m. Fig.6.6 illustrates the average QoE obtained from each scheme. It can be seen that with traditional schedulers (RR, CI, and PF), QoE scores are acceptable only when number of background node is 4. It means that if operators want to ensure QoE at acceptable rate using these schedulers, they can only admit 4 background nodes. On the other hand, they can admit up to 16 and 20 background nodes using QoE-CI and QoE-PF schedulers respectively. In fact, the number of admissible users is an important factor for network operators; since a higher number of users that can be accepted in the network directly imply a higher revenue that can be reached by network operator. Again, we can observe that NRG performs the best in terms of QoE and the worst in terms of throughput presented in Fig.6.7. This is because the goal of NRG is to privilege video users. It gives much more bandwidth to them and not enough to background nodes, which could explain the lowest global throughput seen in the graph. In such a situation, the number of users does not have great impact on perceived quality, thus, with NRG network operator cannot really regulate quality using this QoE factor.

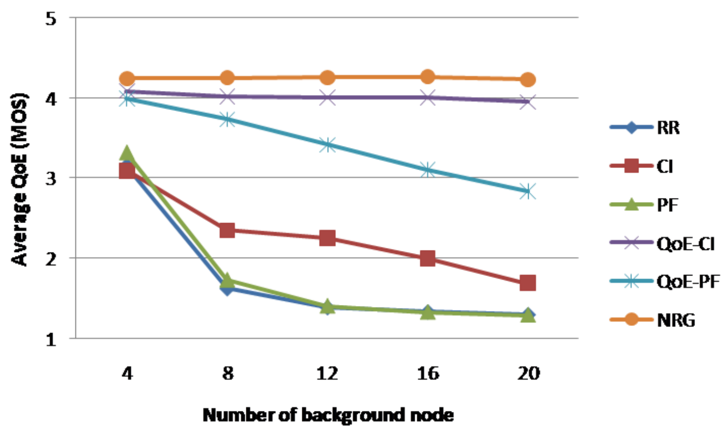


Figure 6.6: Average QoE for different number of background node.

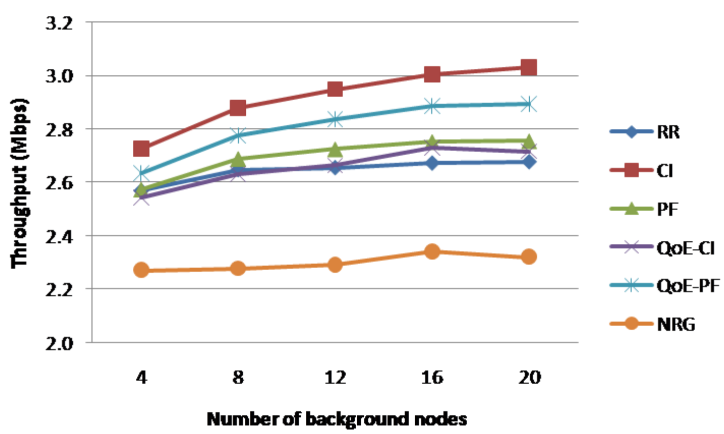


Figure 6.7: Global throughput for different number of background node.

6.6.4 Fairness issue

Two types of fairness are distinguished here: the first one concerns fairness regarding throughput among FTP users and the second one concerns fairness regarding QoE among video users. Consideration of throughput fairness is done only among FTP users as the schedulers will always privilege video users to background ones, thus unfair in this sense. Similarly, QoE fairness is considered only among video users because, for background users, throughput is good enough as quality indicator. Fig.6.8 illustrates the Cumulative Distribution Function (CDF) of per-user throughput, it can be noticed that CI is the most unfair since it gives high throughput (400 kbps) to only 35% of the users and the differences of throughputs obtained from each users are important. We can observe that QoE-CI behaves similarly to CI, thus also a bit unfair. QoE-PF behaves similarly to PF which is proportionally fair. For QoE-PF, about 90% of users get 250 kbps while 80% of users get 300 kbps in PF. It can be noticed that NRG is also fair among FTP users but the throughputs reached by them are much less than other schedulers. Finally, we can observe from 6.9 that traditional schedulers (RR, CI, and PF) are unfair in terms of QoE because about 10% of video users get acceptable score, and the others 90% have to suffer from bad quality. QoE-aware schedulers are fair as well as NRG since all users obtain scores of 3 or higher.

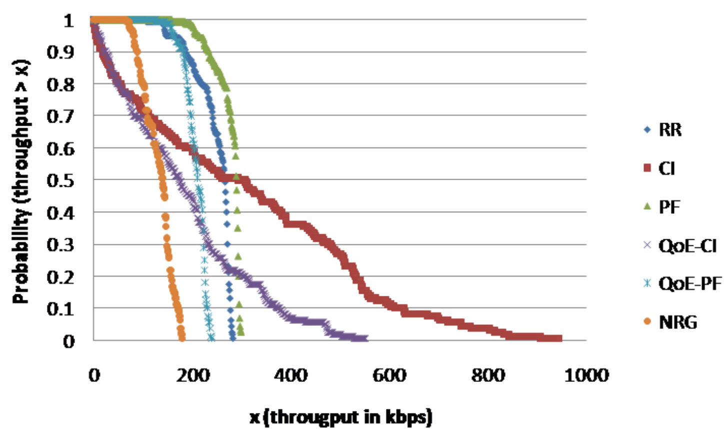


Figure 6.8: Inverse CDF of average per-user throughput.

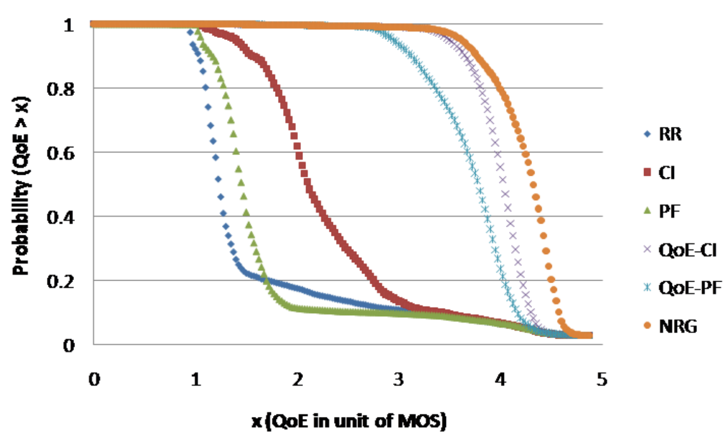


Figure 6.9: Inverse CDF of average per-user QoE.

6.7 Conclusions and Perspectives

In this chapter, two novel schedulers have been proposed for HSDPA; they are aware of ongoing video users' QoE. The proposed schedulers are constructed with the idea to privilege delay-sensitive video users to insensitive background traffic. The results have demonstrated good performance, compromised between traditional schedulers and conservative QoS scheduler. The proposed schemes can be deployed easily on base station providing better control on resources while considering user satisfaction.

To better suit their needs; operators can select the scheduler according to their purpose. If the objective is purely to satisfy video users, then operator may want to use NRG. If operators wish to earn more revenue while keeping video users satisfied, they may want to use one of the QoE-aware schedulers. If they want a fairer scheduler, then operator may want to choose QoE-PF because it takes into account the average throughput of each station and thus fairer than QoE-CI. On the contrary, if operator wants a higher throughput, QoE-CI should be selected because this scheduler privileges station with better signal condition and thus capable of reaching higher throughput.

Furthermore, it can be noticed from the simulation setup that in most of the scenarios the reasonable maximum distance to BS is set to 300 meters. This is to avoid the situation where the station is far away and hence getting a very bad signal quality (CQI). In such case, operators would have to decide if they want to admit or refuse the connection. The idea of admission control in chapter 4 can be applied here knowing that if this type of bad CQI connection is accepted, the quality should also be guaranteed for them.

Another interesting issue for investigation concerns priority differentiation between multimedia traffic classes like VoIP and video streaming. Their characteristics (e.g. traffic patterns, quality requirements, etc.) are different and different treatments would be necessary. Hence, it will be interesting to study how to propose an efficient scheduler to deal with this differentiation as well.

Part III

User-centric Connection Management

We have seen from Part II that using quality of experience in network-centric solutions can make network operator obtain promising results. This Part III will provide investigation on how to deploy quality of experience in mechanism from user perspective or what we call user-centric approach. As can be noticed, there are not many actions that user can take in network management since generally network operator is the one who controls how resources are distributed. However, one of the mechanisms, called network selection, is usually decided by user. Seeing that it is the most studied mechanism in user-centric approach, investigations have been conducted here. In fact, network selection mechanism plays an important role when user needs to choose the best network among available candidates. As terminals nowadays are equipped with multi-interfaces, they also have to choose the network technology that best meets their connection requirement. In this part, different strategies will be presented, among them QoE-based approach is considered as the most relevant regarding user satisfaction. The simulations and results will be illustrated in both homogeneous and heterogeneous wireless environment.

Chapter 7

Network Selection in Wireless Local Area Networks

7.1 Introduction

With increase of multimedia traffic, quality of experience needs to be satisfied at users whilst overall performance needs to be maintained at networks. In order to achieve these goals, the use of network selection mechanism is helpful. When several access points are present, user should select the best available network while trying to keep load balanced between access networks. Therefore, this chapter presents a user-based and network-assisted scheme for *network selection* in wireless LANs. By providing users with relevant information about the network in decision making process, the proposed solution keeps compromising advantage for both user and network operator. The rest of chapter is organized as follow. It first begins with backgrounds and related works in Section 7.2. Then, it continues with description of network selection scheme by detailing functionalities of access points and mobile hosts in Section 7.3. Section 7.4 explains the implementation and results are discussed in section 7.5. Finally, conclusions are given in Section 7.6.

7.2 Problems

Since wireless LANs have started to be deployed, the number of Internet users continue to increase significantly as users can connect easily to the Internet. Nowadays, Wi-Fi hotspots are present everywhere. At the same time, user equipments become more affordable; thus, users with real-time multimedia traffic such as video streaming and VoIP are ubiquitous. This type of user requires specific quality depending on their applications. Moreover, with an increasing number of access points available in the same area, users will have to select the one that will provide the best service for his/her application.

In the standard IEEE 802.11, when a station wants to associate with an existing access point (either after power-up, sleep mode, or just entering the coverage area), it needs to get synchronization information from available access points. This information can be obtained by one of the two methods: 1-*Passive Scanning*: in this case the station only waits to receive Beacon Frames from access points (the beacon frame is a periodic frame sent by the access point with synchronization information); 2-*Active Scanning*: in this case the station tries to find an access point by transmitting Probe Request Frames, and waits for Probe Response from the access point.

After the scan, an association to access points will be decided solely by users. It means that they can connect to any access point they want. A simple decision is usually based only on signal strength measured at the receiver. So in general, users will choose the closest access point because it provides the strongest signal. Using this strategy can sometimes lead to the problem of excessive demand on one access point and underutilization of others. This happens frequently in hotspots as in coffee shops, train stations, or libraries where many users can be found. The new user always selects the access point with the strongest signal without knowing actual load of the network or actual quality experienced by ongoing user. If unfortunately the chosen access point is already high-loaded, one more connection may result in severe degradation of quality for all users of this network. In order to prevent this situation from happening, we need to have a better selection strategy, which provides pertinent information about status of the network, to help users make a good decision.

For that, the IEEE 802.11 Task Group "k" is developing an extension to the IEEE 802.11 standard, referred to as 802.11k [102]. This extension is a specification of radio resource measurement, which is intended to improve the provision of traffic in the physical and medium access layers by defining a series of measurement requests and reports that can be used in selecting the best available access point. Some of the frames are summarized here: *beacon report* (provides information including signal strength and signal to noise ratio), *frame report* (provides information about all received frames), *channel load report* (provides information about busy and free slots), *noise histogram report* (provides the expected value of noise), and *station statistic report* (provides different MAC counters information). However, the objective of IEEE 802.11k is to provide radio resource measurement and not radio resource management. Hence, there is not any decision mechanism defined in this draft.

With considerations stated above, this chapter presents a network selection mechanism by making use of 802.11k concept for communication between networks and users, and by deploying quality assessment tool as support for network selection. Basing on the concept of 802.11k, instead of giving radio measurement information, QoE information is appended into *Beacon* and *Probe Request* frames. The proposed scheme is a *user-based* and *network-assisted* approach. Unlike in other user-based schemes, it does not have problem of load balancing. Indeed, even though users in the scheme select the network by themselves, they take the mean opinion score of overall users into account while making the decision (network-assisted approach). As a consequence, they will connect to the network where they will be best connected and avoid high-loaded networks automatically due to the lower MOS in those networks. Therefore, the scheme is profitable for preventing access networks from over- or under-utilization.

7.3 The Proposed Scheme

This section describes the user-based and network-assisted scheme to solve network selection problem. Users in the scheme can make the decision to which network they will be associated by assistance from access points in the area. Functionalities of access points and users are described respectively. The quality-affecting parameters chosen in this scheme are loss rate (LR) of I/P/B and mean loss burst size (MLBS) of I frame, as previously described in chapter 5 (5.4.1).

7.3.1 Access Points Functionality

To avoid the situation while users lack pertinent information to make decision, the access point in this scheme sends current QoE perceived by ongoing connections to new comers so that they can decide to connect to the best available network. Indeed, the sending information is the average of mean opinion score of all ongoing users at the access point. This can be achieved by embedding MOS into Beacon and Probe Response frames. When passive users receive beacons, they will also receive MOS of all presenting networks. Similarly, when active users send Probe Request, they will receive Probe Response along with the corresponding MOS. With this information, users can make the best decision by choosing the network that has the best QoE condition.

It is assumed that access points in the scheme operating the same way as in chapter 5 for feedback mechanism; this means the AP sends out request for MOS from users who return it back afterwards. The period of computing new MOS (average MOS of the network) should not be less than beacon's broadcasting interval since access points will broadcast the MOS in beacon frames. In addition, it should not exceed

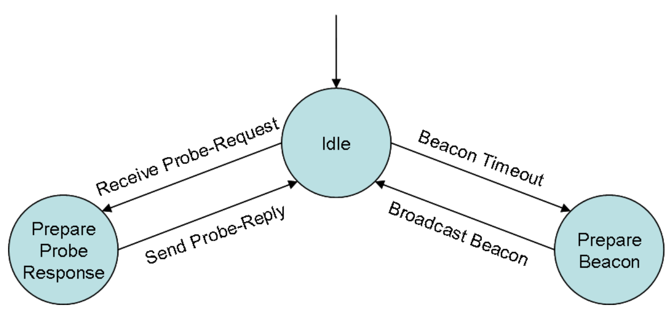


Figure 7.1: Access Point States.

user arrival rate, otherwise the sending MOS will be out of date. The most appropriate interval should be approximately equal to the user arriving interval and the computation should be done after acceptance of new connection. The automaton of access points is depicted in Fig. 7.1, which presents only the three states of access points that concern the scheme: *Idle*, *Prepare Beacon*, and *Prepare Probe Response*. When the Beacon timer rings or when the Probe Request is received, the access point prepares the frame and broadcasts MOS within Beacon frame and Probe Response frame respectively.

7.3.2 Mobile Host Functionality

The context is network environment with only one network operator, this reduces a complexity resulting from different prices charged by different network operators. Since only one network operator is assumed, the prices of all access networks are assumed the same. The investigation about the effect of pricing can be done furthermore to treat the case where more than one network operators are present and the scheme can be refined accordingly.

In the proposed scheme, users select the network that provides the highest score or they may not connect to any access point if they consider that the current scores are too low for the requirement of their applications. The minimum requirement can be

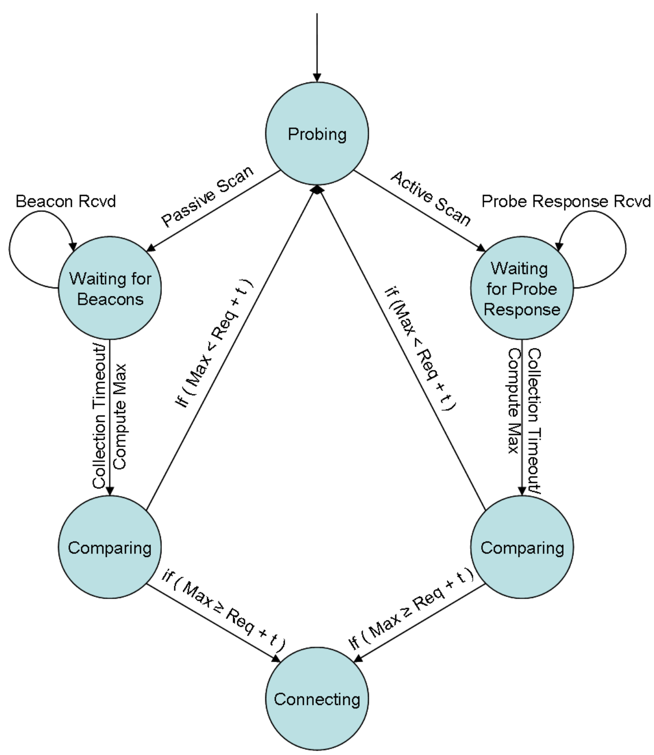


Figure 7.2: Mobile Host States.

regulated by user. Indeed, the user will have to compare the maximum of the receiving MOS (*max*) with the required score (*req*) plus a threshold (t) that corresponds to the degradation margin of the network after acceptance of the new connection. If the maximum MOS is higher, then user will request for connection from the access point whose *max* belongs to. Otherwise, it will not request for connection since the minimum QoE for the application cannot be satisfied. The mobile host automaton is depicted in Fig. 7.2.

It can be noticed that the threshold t is very delicate to define as it depends on the granularity expected by the application. If t is high, it will result in high quality because the scheme will restrict the selection to the network that has high degradation margin of MOS, this type of network grants all its capacity to a small number of users who greatly benefit from it. However, this restriction raises underutilization problem to the network in the case that none of candidate networks satisfies user requirement and the available bandwidth is not allocated to anyone. With the similar reasoning, if t is small, it will be more vulnerable to quality degradation when the number of traffic increases, thus congestion resulted from the new comer in the network. Therefore, a tradeoff between bandwidth utilization and its consequence in connection degradation has to be well investigated.

Except few rare cases that user fixes the requirement extremely high or extremely low, with this user-based and network-assisted scheme, the problem of overloaded or under-utilized networks are solved automatically because users will select the network that has the highest MOS (under-utilized networks) and avoid the one that has smaller MOS (high-loaded networks).

7.4 Performance Evaluation

This section gives explanations of the scenario, the implementation, and the simulation of the proposed scheme comparing to those of signal-based schemes. A simple scenario is considered, it consists of one type of traffic (video streaming) requested by all users.

7.4.1 Simulation setup

All mobile hosts in this example want to watch video streaming on the mobile device. The video specification is an H.264-coded sequence of duration 64 seconds and encoding rate about 384 kbps. The video data can be streamed from a video server on the wired network to the terminal through different access points. At the beginning of every second, a new station is asking for a connection, which means the connection arrival rate is one connection per second. For this example, MOS equals to 3 is considered enough for video streaming application, the station chooses the access point that has the highest MOS and verify that the MOS is at least equal to $3+t$. In case of multi-operator scenarios, user may decide to choose the network that has at least $3+t$ with lower price; it is not necessary to choose the access point with the highest MOS.

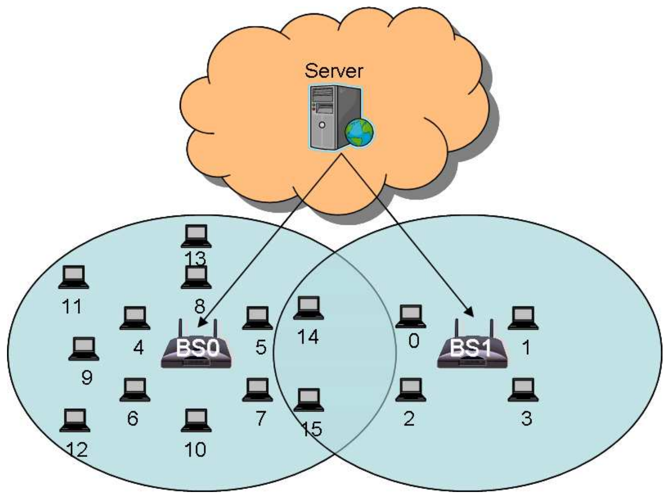


Figure 7.3: Network topology in the example.

The users are faced with the scenario like the one depicted in Fig. 7.3, where the decision of which access network to use for transporting the video streaming application is required. The topology consists of two access points (BS0 and BS1). There are a total of 16 stations, each requests for connection one after another according to the station ID, meaning that station 0 (ST0) begins to ask for connection first and then station 1 (ST1) and so on. Coverage areas of the two access points are illustrated with its corresponding circle, therefore station 14 (ST14) and station 15(ST15) have possibility to connect to either BS0 or BS1 due to the overlapping coverage. The other stations are situated in only one coverage area either the one of BS1 (station 0 to 3) or BS0 (station 4 to 13). We can see that ST14 and ST15 are closer to BS0 and thus receive higher signal strength from this access point. In general case, both stations will automatically choose access point BS0 that provides the highest signal strength and make the network of BS0 overloaded. On the contrary, with the proposed scheme, every time a station has to make a decision, it always selects the network that has the highest MOS instead of the one that has the highest signal strength. This help solving load balancing issue at the same time.

7.4.2 Implementation of network selection mechanism in WLANs

For the implementation, the scheme takes the threshold $t = 1$ after extensive simulations and it is the reasonable value to protect overall quality. Simulations are done by the network simulator NS-2 [21] version 2.29 with the wireless update patch from [83] with improvement from original support as described in 5. Communications between users and access points for the feedback procedure refers to the use of IEEE 802.11k standard. Since the frames in this draft have available fields, MOS is put in one of them. While broadcasting Beacon or responding to Probe Request, the access point informs users about MOS at the same time. The user receives and extracts MOS from all presenting access points and selects the network that has the highest MOS. If only one access point is present, the user can decide whether to connect to the access point or not, based on the receiving score and application requirement.

7.5 Results

This section presents the result of the scheme based on MOS comparing to the one based on signal strength. The satisfactions of both users and networks are necessary. User satisfaction is considered in terms of: individual MOS achieved by each user, global MOS (to see the overall satisfaction of users in the access network), and fairness (to see if MOS is fairly distributed among users). For the network, satisfaction is considered in terms of load distribution.

7.5.1 User satisfaction

Fig. 7.4 illustrates user individual satisfaction. This graph depicts the satisfaction in terms of QoE obtained by each scheme, which is the quality of experience perceived by users. Note that, stations with ID. 0 to 3 are the ongoing connections on BS1 and those with ID. 4 to 13 are the ones of BS0, and ST14 and ST15 are the most recent comers that are located in the overlapping coverage of the two access points but closer to BS0 than BS1. The decision to be made is which access point ST14 and ST15 are going to request for connection.

In a signal-based scheme, ST14 and ST15 will choose BS0. On the contrary, when applying QoE, both ST14 and ST15 will find out that MOS in BS0 is lower than in BS1, and they will connect to the BS1 instead. The results obtained in Fig. 7.4 have shown that QoE-based scheme outperforms the one based on signal strength. We observe the difference of MOS as high as 3 levels (QoE improvement from *Poor* to *Excellent* for ST14 and ST15). All stations of BS0 in QoE-based scheme show profitable increase in quality as well.

Moreover, the results of average MOS in network of BS0 and BS1 are also illustrated in Fig. 7.5 and in Fig. 7.6 respectively. We observe a better performance of average MOS in the network of BS0; however, the averages of MOS in BS1 of both schemes are the same because the network is low-loaded and can provide high quality of service to all users. It can be noticed that the selection scheme is efficient when the network reaches a certain load (10 connections in this case); before arriving to this point, it will not reveal benefit as in the case of Fig. 7.6. Both figures also present overall MOS after all stations are connected until the end of transmissions.

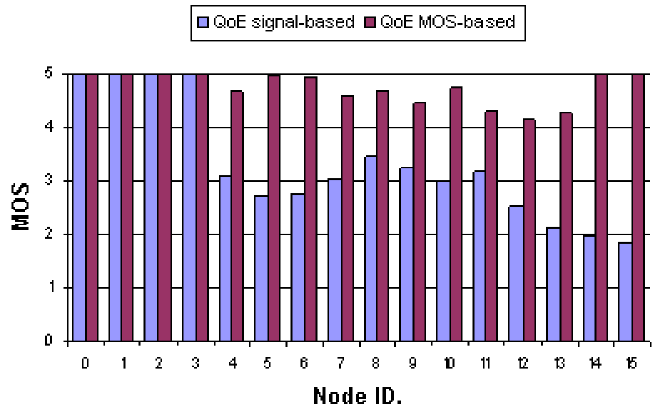


Figure 7.4: Individual User Mean Opinion Score in BS0 and BS1.

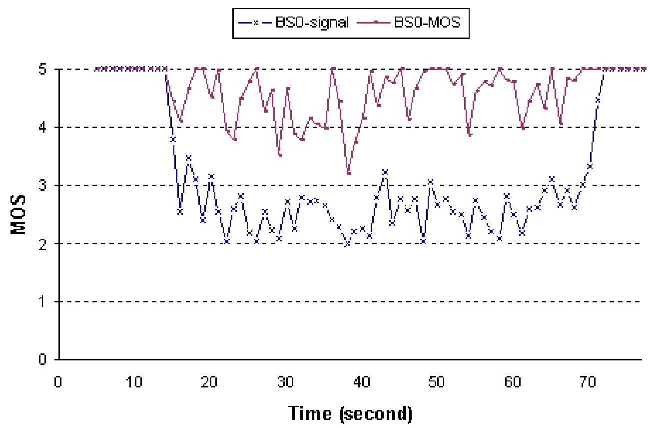


Figure 7.5: Overall User Mean Opinion Score in BS0.

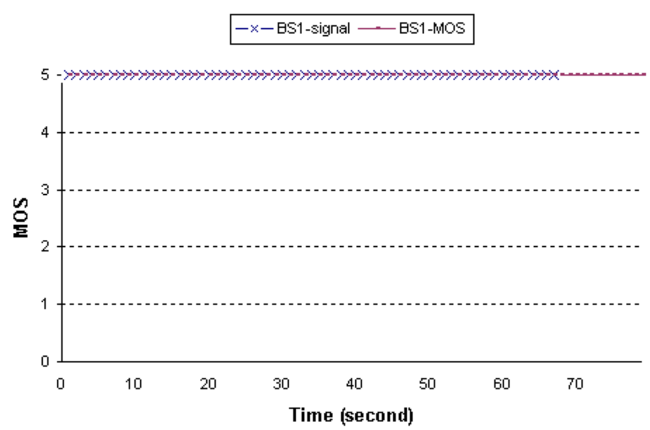


Figure 7.6: Overall User Mean Opinion Score in BS1.

7.5.2 Load Balancing

Fig. 7.7 illustrates loads of each access network. The y axis represents the load in terms of number of connections and bandwidth utilization that can be computed approximately by this number (n) times video bit rate ($n*384kbps$). It can be noticed that QoE-based scheme performs better in terms of load balancing between the two access networks. The difference between loads of the network representing by BS0 and the one of BS1 in this scheme is significantly smaller than the difference of signal based mechanisms. This is automatically obtained with user selections since they will prefer the network with higher MOS and usually low loaded to one with lower MOS, generally high-loaded.

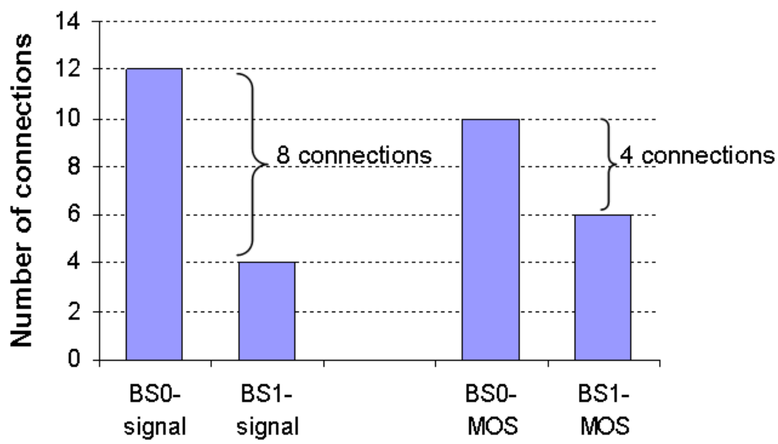


Figure 7.7: Load distribution among access networks.

In general, load balancing problem are usually solved by limiting number of connection or requested bandwidth, which leads to a very conservative approach. Here, MOS is considered instead of those parameters because sometime one connection may require more bandwidth than the other in order to have satisfied QoE; for example, the video that has lots of movement will require more bandwidth than the one with less movement. Therefore, using QoE is more appropriate and more flexible.

7.5.3 Fairness Index

Fairness in terms of MOS is computed to see whether QoE is fairly partitioned or not among users. Jain's Fairness Index [111] is used for the computation of the network of BS0, the one of BS1, and the overall network, as the following equation:

$$f(x_1, x_2, x_3, \ldots, x_n) = \frac{(\sum_{i=1}^{n} x_i)^2}{n \times \sum_{i=1}^{n} x_i^2}$$

where x_i is the MOS obtained at the station i and n is the total number of stations in the network. With $0 \leq f() \leq 1$, the more the output of this function close to 1 the better fairness is partitioned.

The result in Table 7.1 demonstrates that the proposed scheme is fairer than signal-based scheme. The fairness index obtained for the global network (including BS0 and BS1) obtained in QoE-based scheme is as high as 0.996 while that of signal-based scheme only reaches 0.905. Similarly, in BS0 we obtain better index, 0.997 comparing to 0.968. Nevertheless, the indexes of both schemes in BS1 reach 1 as they both satisfy every user equally in case of few users. This can be explained by the fact that the small number of users can profit from the whole provided bandwidth in both cases.

Table 7.1: Fairness Index between signal-based and MOS-based schemes.

Scheme	Global	BS0	BS1
Signal-based	0.905	0.968	1
MOS-based	0.996	0.997	1

7.6 Conclusions and Perspectives

This chapter presents QoE-aware network selection scheme, a novel technique to handle network selection problem for multimedia users. In this scheme, user selects the network that has the best QoE (seen from connected users). The scheme is user-based and network-assisted, which is profitable for both user and network operator. The results have illustrated that it performs well and gives high QoE for new and ongoing connections. Moreover, load distribution is well balanced.

Further, the similar strategy can also be applied to other wireless technologies such as WiMAX and Cellular networks. In the following chapter, several network technologies will be combined together to constitute a heterogeneous environment, and a strategy using the similar concept will be defined. Utilization of this idea is particularly helpful in heterogeneous networks because, unlike the technical parameters, MOS is technology-independent, which make it applicable to all technologies.

Chapter 8

Network Selection in Heterogeneous Wireless Networks

8.1 Introduction

Deployment of next-generation network (4G) begins to spread throughout the world. With variety of network technologies, it is possible for users to select an appropriate network that best suits their needs. The problem is how to provide the mechanism that helps users in making decisions under heterogeneous environment. Even though many schemes have been proposed in the literature but none of them takes into account QoE. As it represents perception experienced by the user, it is thus an essential indicator for 4G networks, especially with multimedia communications nowadays. Therefore, this chapter presents a novel network selection mechanism that takes quality of experience into consideration for decision making. Similar to the previous chapter, it is a user-based and network-assisted approach thus a compromise solution between user and network benefit. The main idea is to include MOS of ongoing users in candidate networks as one of indicators to select the best network for connection. The rest of this chapter is organized as follow. Section 8.2 gives a comprehensive survey of related works that presents recent schemes having as objective the network selection in heterogeneous environment. The chapter continues with the proposed scheme in section 8.3. Then, test setup is described in section 8.4 and the results are presented in section 8.5. Finally, conclusions and open directions are given in section 8.6.

8.2 Related Works

The emergence of heterogeneous network has pushed the research in this area to progress very rapidly and many schemes have been proposed. The related works will refer to part of the survey in chapter 2, the concerning works are summarized here.

The authors of [41] have proposed *Customer Surplus* function to deal with non real-time transmission. In this protocol, users first survey their network interfaces and determine the list of available access networks. Next, they predict the transfer rate of each available network taking the average of the last five data transfers and then derive completion times. After that, they compute predicted utility, which is the relationship between the budget and the user's flexibility in the transfer completion time. Finally, for each candidate network, users compute consumer surplus, which is the difference between utility and cost charged by the network and they choose the best one to request for connection. It can be noticed that this scheme works fine in non real-time traffic but not for real-time multimedia service that is the most popular nowadays.

To handle handoff, the authors of [51] have proposed *Profit Function*. The authors associated each handoff with a profit that is decided by a target function with two parameters: *bandwidth gain* and *handoff cost*. Parameters used in the calculation of the gain include: (i) access networks along with their maximum bandwidth provided to a single user as well as capacity utilization; (ii) application's maximum requirement on bandwidth; (iii) access networks' bandwidths used by a mobile node for handoff. Then the authors defined a handoff cost as data volume lost due to handoff delay; it corresponds to the volume of data which could have been transmitted during the handoff delay. Thus, the profit is a difference between gain and cost. At each handoff epoch, mobile node compares profit from each network and chooses the one that yields maximum profit. This scheme takes only bandwidth-related parameters into account. However, considering solely bandwidth cannot guarantee good QoE for multimedia applications.

The authors of [43] have proposed network selection using *analytical hierarchy process* to weigh QoS factors and using *grey relational analysis* to rank networks. With QoS factors, the authors constructed an AHP hierarchy based on their relationships. QoS is placed in the topmost level as the objective; main QoS factors describing network conditions are placed in the second level. Moreover, factors have been decomposed into sub factors and they have been arranged in the third level. Finally, available solutions are arranged in the bottommost level. User-based data is collected and processed by AHP for weight computation. At the same time, network-based data are normalized by GRA, and then ideal network performance is defined following by calculation of the grey relational coefficient which gives grey relationship between ideal network and the other. The calculation of GRC takes the previously computed weights into account; finally, the network with the largest GRC is the most desirable. This scheme takes many technical parameters into account but still does not include QoE, an essential factor for multimedia users.

Also deploying multi-attribute decision making, the authors of [52] have proposed an algorithm based on *Fuzzy Logic Controller* to evaluate fitness ranking of candidate networks. They differentiate decision making into three phases: pre-selection, discovery, and decision making. Pre-selection phase takes criteria from user, application, and network to eliminate unsuitable access networks from further selection. The authors implemented discovery phase based on fuzzy logic control, they fuzzify crisp values of the variables (network data rate, SNR, and application requirement data rate) into grade of membership in fuzzy set. Then these membership functions are used as input to the pre-defined logic rule base. Finally, overall ranking is obtained through defuzzification with weighted average method. It needs to be mentioned here that fuzzy logic control gives good result in this case of few metrics. However, if the metrics number increases, the system may become very complex and may give erroneous results.

Even though all proposed schemes have covered many aspects and have taken into account several parameters, they cannot guarantee users' satisfaction since none of them is interested in quality of experience metric, which is the most prominent factor in heterogeneous networking today. Therefore, the QoE-based mechanism has also been studied in heterogeneous environment. For a better comprehension, Table 8.1 presents each scheme and its corresponding parameters.

Table 8.1: Different Network Selection Approaches.

Scheme	**Parameters**	**Nature of Parameters**
Customer Surplus	Transfer rate and cost	Technical
Profit function	Available and required bandwidth	Technical
AHP & GRA	User requirements and network conditions	Technical
Fuzzy Logic	Network and application data rate, SNR	Technical
This proposition	*Quality of experience*	*Subjective*

8.3 The Proposed Scheme

This section describes the decision mechanism then it gives example of scenario that will be used for the test. To provide information to users for decision making, a point of attachment in this scheme broadcasts QoE information to all users within its range. The embedded MOS is the minimum score among all ongoing users of this PoA or perfect score if there is no ongoing user. The minimum score is diffused because the mobile node should be aware of what the worst quality it can get after the connection request. This can be done via signaling messages in IEEE 802.21 MIH (media independent handover) [31].

Let OF be the objective function to be computed for each network. It is calculated by the sum of each criterion i (C_i) times their weight (w_{ci}). Weight can be set as desired by users (all weights are equal by default). Assuming n represents the number of criteria, OF can be written as in equation (1) below.

$$OF = \sum_{i=1}^{n} C_i * w_{ci}; \text{ where } \sum_{i=1}^{n} w_{ci} = 100 \qquad (1)$$

The value of C_i is then normalized by the maximum value, which gives C_i a value in the range [0..1]. The sum of all weights is equal to 100, thus the score of each network is in a range of [0..100]. After having computed OF for all available networks, the mechanism selects the network that has the highest score for requesting connection. The other networks are arranged in a ranking table. If the connection request of the first choice network cannot be satisfied by network operator, the station tries the next one in the table respectively.

Taking an example, it is assumed that a mobile node (MN) is multi-mode; it is equipped with Ethernet, WLAN and 3G interfaces. Major factors influencing user decisions in network/handover selection are quality of experience (*qoe*), cost (*cost*), and mobility (*mob*). By default, raw values of each criteria are in the range [1..5]; hence, the OF of network technology k can be written as equation (2) below.

$$OF(k) = C_{qoe}(k) * w_{qoe}(k) + C_{cost}(k) * w_{cost}(k) + C_{mob}(k) * w_{mob}(k) \quad (2)$$

Table 8.2 presents an example of criteria scoring. It can be noticed that QoE is the only parameter to be measured; the other two can be taken directly from the table.

Table 8.2: Example of Criteria Scoring.

Technology	**Quality of Experience**	**Cost**	**Mobility**
Ethernet	to be measured $(x/5)$	free $(5/5)$	none $(1/5)$
WLAN	to be measured $(y/5)$	low $(3/5)$	low $(3/5)$
UMTS	to be measured $(z/5)$	high $(1/5)$	high $(5/5)$

To have some guarantees on QoE, *threshold-based* mechanism is proposed; the threshold indicates a border beyond which the quality of experience may not be guaranteed. This step is done after network ranking to ensure that the winning network can suit user satisfaction. For that, the mobile user sets its threshold MOS (mos_{th}) then compares it with minimum score (mos_{min}) obtained from the winning network. This threshold is defined as the acceptable MOS plus an absorber; i.e. $mos_{th} = mos_{acpt} + mos_{abs}$. If the minimum score is higher or equal to this threshold, then the connection request is launched. Otherwise, the mobile node may revise its weight assignment or QoE expectation. One exception exists, in which we called it *forced handover*. The connection request is launched even when the minimum score is less than threshold. For this case, when the candidate network is the only available network in the area; if handover is not executed, the mobile node will lose its connectivity. Please note that the absorber is very delicate to define as we deal with quality of experience. To ensure high quality of experience, user may set this absorber to a high value but the trade-off is that it may not find an appropriate candidate if the expectation is too high.

8.4 Performance Evaluation

The proposition is compared with a scheme, called *Priority-based*, in which the decision making is based on priority classification. This priority concerns network interface technology/type. The highest priority goes to Ethernet interface, following by WLAN, and UMTS technology respectively. This classification is implemented in real Mobile IP tool such as Segco Mobile IP [112] as well as in NS-2 from NIST [23]. The reason for this classification is very high bandwidth and no cost of Ethernet, following by medium bandwidth and low cost of WLAN, and low bandwidth and high cost of UMTS regardless of its high mobility. This section first describes the implementation and test setup along with the testbed configuration and topology, and then it continues with the obtained results.

8.4.1 Implementation of network selection mechanism in HWNs

The implementation is based on NS-2 with NIST add-on [23] (mobility extension: IEEE 802.21 model and 802.11), which enables simulations of heterogeneous environments. This simulation platform incorporates a variety of access networking technologies to run jointly. In the original `handover` module from NIST, handover selection is done according to priority. This means, a terminal connects to a new network if it is better than the current one according to the order of technology. For the tests, this module is modified in order to add the decision making based on quality of experience as previously described.

8.4.2 Simulation Setup

The scenario is presented in Fig.8.1. Mobile node (MN) is a multi-interface terminal. It is equipped with UMTS and WLAN interfaces. At the beginning, the only available network present is UMTS so the MN starts its connection via UMTS. The MN moves during the connection until it enters WLAN coverage (after 24s). There are two possibilities, either MN stays in the same network or MN hands over to WLAN.

The scheme deploys $mos_{acpt} = 3$ because this value is the standard acceptable level of QoE for video streaming application. As for mos_{abs}, tests have been conducted with different values (2.0, 1.5, 1.0, 0.5, 0.0) in order to see how network behaves. Fig.8.2 shows how user experience and global throughput behave with decreasing values of absorber. Please note that the throughput here is considered in terms of accepted number of flows in the system; this is to see how network operator admits traffic with different values of absorber. As mentioned before, if this absorber is high and an appropriate network exists then quality of experience will be very good. However, if we analyze closer we can see that throughput in this case is very low; the reason is because network dedicates the whole bandwidth to only a few connections. In addition, as the expecta-

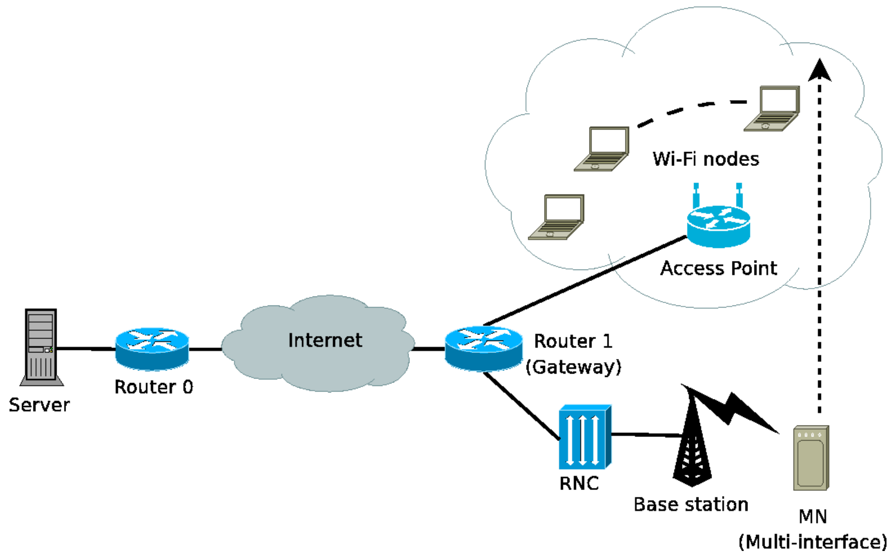


Figure 8.1: Network Topology.

tion is high; hence, it is difficult to find an appropriate network. On the contrary, if the value of absorber is lower, the quality of experience decreases and the global throughput of the system increases accordingly. Considering all criteria mentioned above, the scheme deploys $mos_{abs} = 1.0$ and thus $mos_{th} = mos_{acpt} + mos_{abs} = 4$.

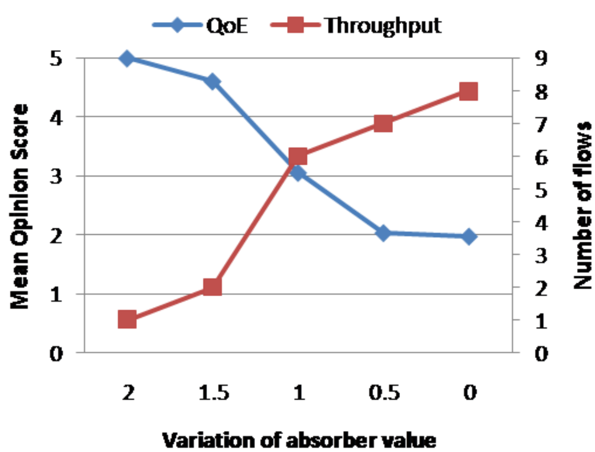


Figure 8.2: Network behavior with different absorbers.

Two scenarios are investigated: scenario 1 with moderate-load condition and scenario 2 with high-load condition. They will be used to demonstrate that if everything is doing fine, no precaution or management mechanism are needed. However, when the condition degrades, some adaptations needs to be done in order to alleviate the situation. This section will show how the proposed mechanism can guarantee mobile node having good quality of experience. It also provides the preliminary result for introducing admission control, which can be done by network operator to also ensure quality of ongoing users.

8.5 Results

In this section, results from the previously described scenario are presented in terms of quality of experience (MOS) and bandwidth utilization (throughput).

8.5.1 Moderate-load condition

The most important metric is user satisfaction. For measuring user satisfaction of the running application, the quality of experience is considered in terms of MOS as previously described. Fig.8.3 presents the quality of experience perceived by MN. We can clearly observe perfect scores obtained with QoE-based scheme. On the other hand, if MN decides to hand over to WLAN, the quality will slightly fluctuate during connection holding time. Regarding the quality of experience obtained by ongoing connections within the WLAN. The graph in 8.4 presents the lowest scores among all WLAN users in time. It can be noticed that QoE-based scheme performs slightly better than the priority-based scheme but there is not much difference. Nevertheless, minimum scores obtained with QoE-based scheme stays above 4 (*Good* quality) most of the time and does not decrease below 3 (*Fair* quality). On the contrary, scores obtained with priority-based scheme go below 3 (*Fair* quality) and reaches 2 (*Poor* quality) twice. Since there is not any other traffic in UMTS, MN would rather stay in the same network where it could get perfect quality than hand over to WLAN where quality is fluctuating. However, the fluctuation in this case is not crucial as it stays above 4 all the time.

Fig.8.5 and Fig.8.6 present bandwidth utilization in UMTS and WLAN respectively. It can be seen that QoE-based scheme provides a better balance of load between the two networks. This is because load is automatically distributed by MOS indicator. User selects network with higher MOS, which is generally low-loaded, and hence load is better distributed. On the contrary, when using priority-based, the scheme does not take any concern of quality into account and blindly change user into WLAN expecting larger bandwidth and lower cost.

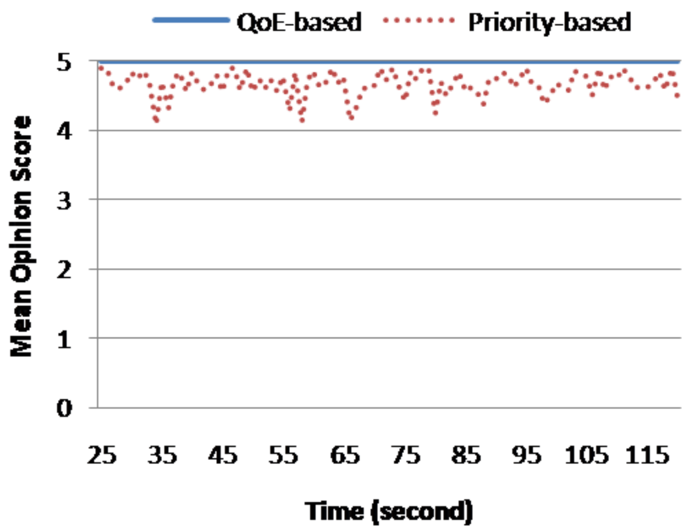


Figure 8.3: Quality experienced by MN under moderate load.

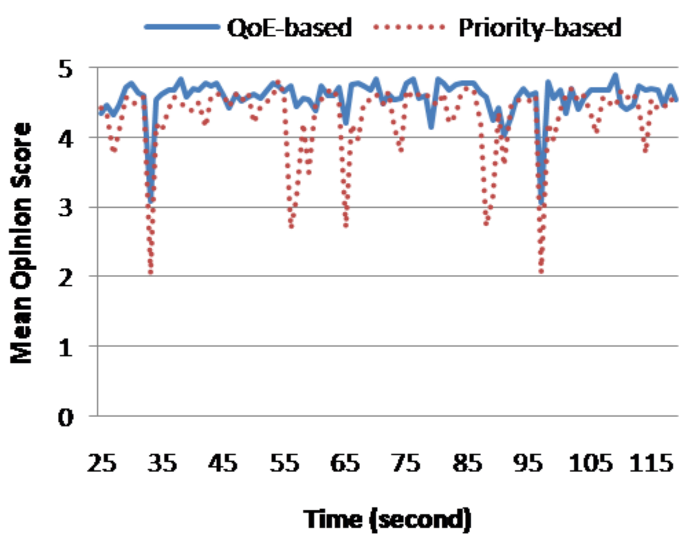


Figure 8.4: Quality experienced by WLAN nodes under moderate load.

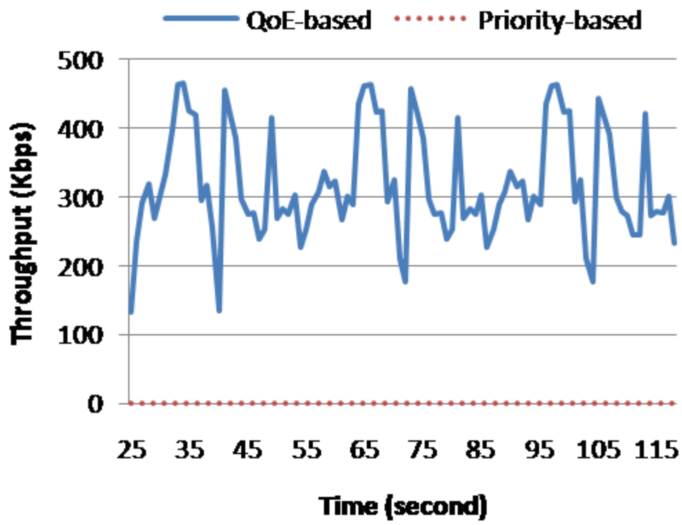


Figure 8.5: Throughput in UMTS network.

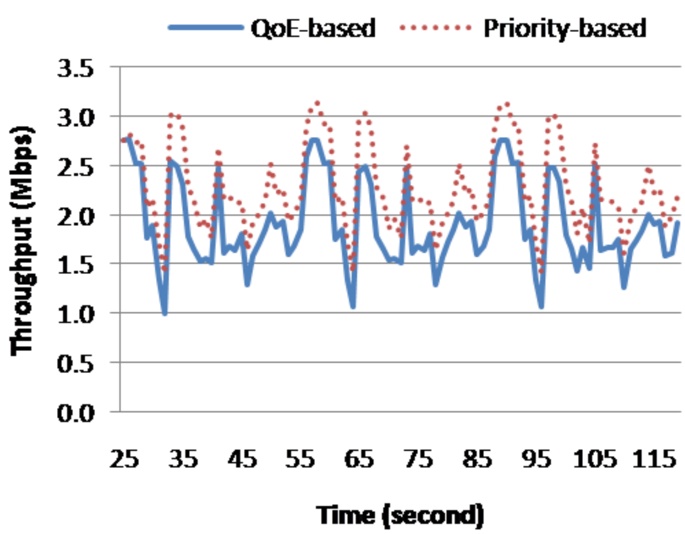


Figure 8.6: Throughput in WLAN with moderate load.

8.5.2 High-load condition

In order to show how the situation can become much worst, this scenario illustrates the case when WLAN is high-loaded. A new user enters to the network every second and hence increasing load in time (connection holding time is 60 seconds). The MN decides whether to execute or not a handover in this situation.

Fig.8.7 presents perceptual quality experienced by MN. The blue curve results from QoE-based scheme, in which the MN decided not to enter WLAN after seeing MOS condition of ongoing users. The red curve results from priority-based scheme, in which the MN continues to make a handover to WLAN regardless of current WLAN condition. We can observe great improvement as MN obtains perfect scores along the session with our QoE-based mechanism. On the contrary, it obtains a very fluctuating score with priority-based scheme and sometimes quality drops closed to 1 (*Bad* quality). As for ongoing users in WLAN, our scheme outperforms priority-based scheme by providing good quality of experience, minimum MOS is close to 5 (*Excellent* quality) most of the time. On the other hand, minimum MOS of priority-based scheme performs badly. Even though, majority of score is above 3 (*Fair* quality) but it drops close to 2 (*Poor* quality) several times.

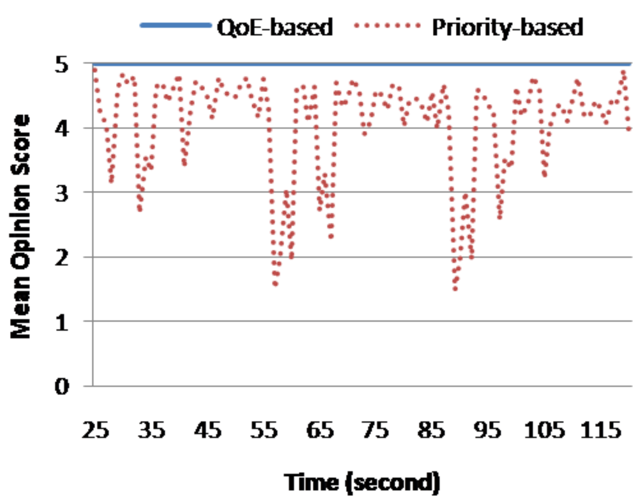


Figure 8.7: Quality experienced by MN under critical condition.

For bandwidth utilization, the result of UMTS load distribution is similar to Fig.8.5 as the scheme leaves the UMTS network with no previous traffic. On the other hand, the WLAN throughput of priority-based scheme is shifted up a little as can be seen in Fig.8.9. This is because WLAN has more traffic flows in the network. It can be remarked here that there is always a trade-off between bandwidth utilization in a network, load balancing between different networks, and quality of experience. In general, network operator wants to take the most profit from available bandwidth and sometimes ignores the result in quality experienced by users. We can see from this

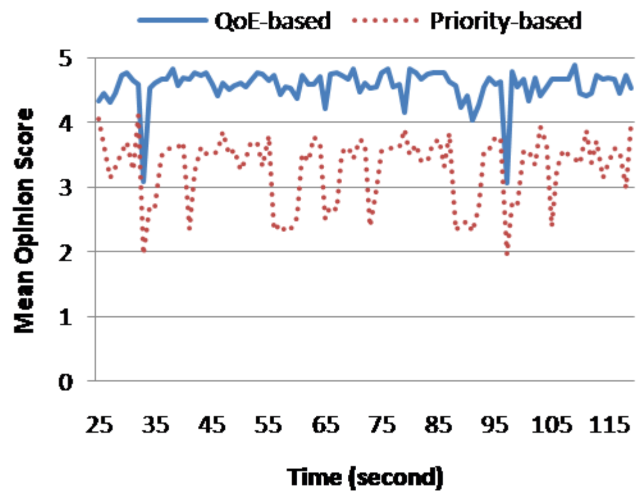


Figure 8.8: Quality experienced by WLAN nodes under critical condition.

example (red curves) that when bandwidth utilization is high in WLAN (Fig.8.9), the QoE of ongoing users becomes poorer (Fig.8.8 and Fig.8.7).

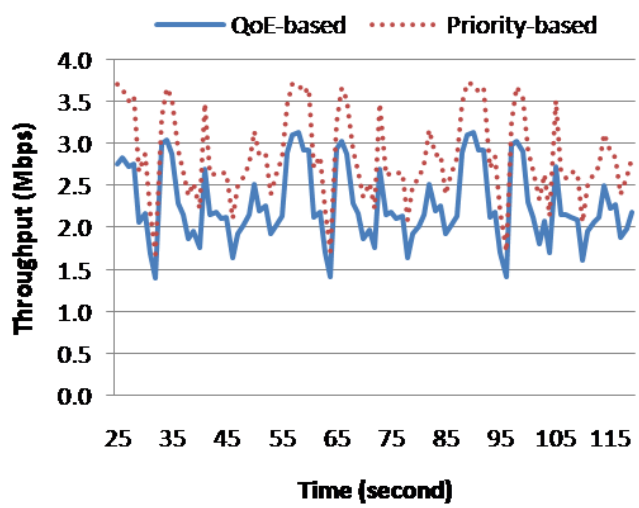


Figure 8.9: Throughput in WLAN under critical condition.

8.5.3 Discussion

It can be seen that network selection mechanism is useful for making decision when entering the network. However, it should be mentioned here that this procedure only guarantee the entrance phase. A bad result can still be obtained later even with a good network selection mechanism. This is the case in which WLAN load continues to increase after the handover of MN. In such a case, quality of experience can continue to degrade until very bad performance. If there is no other network to hand over to, user will have to suffer from this bad situation.

To understand this scenario, deeper investigation is conducted to see how the quality of experience can be influenced by network load. Fig.8.10 presents MOS evaluation with increasing number of traffic in WLAN. The blue curve presents average MOS in time whereas the red curve presents the lowest MOS in time. It can be seen that MOS decreases when network load increases. In this situation, network operator needs to take an action in order to maintain quality of experience at acceptable level. Management mechanism such as admission control can be used for that. For example, the network operator can filter incoming connection with MOS of ongoing users. This can help in maintaining good user experience for everyone.

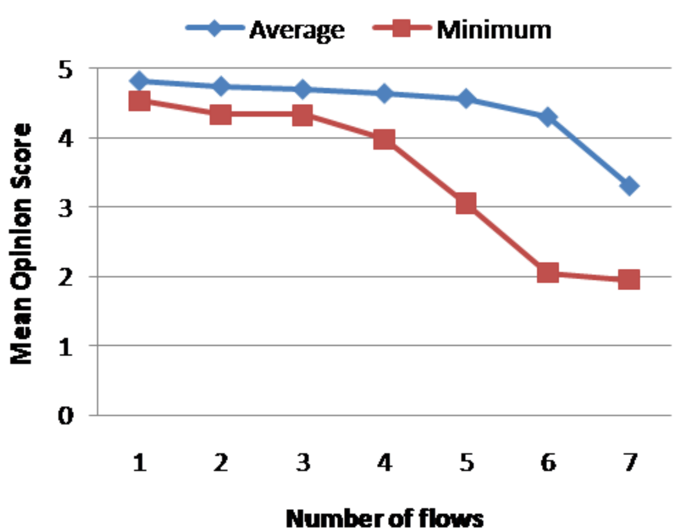


Figure 8.10: Quality experienced by WLAN nodes with increasing traffic.

8.6 Conclusions and Perspectives

In this chapter, another network selection mechanism based on quality of experience has been proposed. The scheme considers different criteria including user experience for making decision. It is compared with priority-based scheme currently in use on many Mobile IP implementations. The obtained results show that the proposed scheme performs better in guaranteeing both quality of handover user (MN) and ongoing users in the target network. Its load distribution is also better as UMTS network can gain some throughputs from the MN.

The obtained results show that even with simple mechanism, we can see performance improvement. Enhancement can further be done using more sophisticate mechanism, for example, multi-attribute decision making with QoE as one of attributes. In addition, it would be interesting to investigate more complex scenario and to compare QoE-based scheme with other handover schemes such as QoS handovers as well.

As we can see from discussion, network selection alone is not enough. It only helps mobile user to select the best network at the moment of connection request but it cannot guarantee that network condition will not change after the selection process is finished; especially, if network condition degrades and no other network exists for the handover. The chapter then gives the primary result of quality evaluation with increasing number of flows, which shows that admission control is also necessary in order to maintain good QoE along connection holding time. The idea of admission control from chapter 4 can be applied, in combination with network selection and under heterogeneous environment.

Chapter 9

Conclusions and Perspectives

9.1 General conclusions

This document provides a thorough investigation of resource management using quality of experience, a new concept of quality that has been recently emerged in multimedia networking today. An appropriate assessment method (PSQA) has been chosen in order to measure QoE in real time. Using statistic learning with random neural network, this method derives user experience using information from the network traffic in real-time manner. With this automatic measurement of QoE, many management directions have been explored and studies have been conducted. This includes management from both network and user perspectives.

The mechanisms concerning network operator are admission control, rate adaptation, and bandwidth scheduling. QoE indicator is used for all mechanisms. In chapter 4, IEEE 802.11 access point admits or refuses new connection according to QoE of ongoing users. In chapter 5, it also adapts multicast transmission rate according to QoE of multicast clients. Moreover, as for scheduling bandwidth in HSDPA (chapter 6), base station can also deploy QoE of multimedia users to prioritize these users to background users and allocate more bandwidth to them in order to satisfy quality requirements. Concerning user side, connection management such as network selection mechanism has been studied using QoE of ongoing users in the candidate networks (chapter 7 and 8).

Investigations have begun in homogeneous environment such as WLAN and UMTS independently and then in heterogeneous environment composing of both technologies. The obtained results illustrate encouraging performance in terms of user satisfaction, bandwidth utilization, load balancing, and fairness, for deployment of QoE as metric in resource management; hence, the objectives of this thesis. It can be noticed that only video streaming application have been studied here; however, the same management ideas can be further applied to other types of multimedia traffic as well. In addition, as QoE is context-independent, it can also be deployed in other network technologies or architectures as well.

9.2 Resource Management with QoE

For a better comprehension, this section discusses limitations and remarks concerning the use of QoE as metric in wireless multimedia network management. As PSQA has been deployed for QoE measurement, its remarks and limitations are also discussed.

First of all, we can notice that even though methods and tools exist for measuring quality of experience with good precision, however they are only used for the assessment purpose. It is very difficult to predict quality of experience that user will perceive in advance (QoE provisioning). For example, in this version of PSQA tool, the statistics used are loss rate and mean loss burst size at packet and application frame level. They are difficult to predict in advance, especially the mean loss burst size. However, using probabilistic or statistic models would be helpful in provisioning these parameters and thus in computation of QoE (via RNN). Therefore, this is a very challenging issue to be investigated in future work.

Another point concerns the quality of experience, which is guaranteed to end users. Network operator need to consider if the providing service is guaranteed with average QoE or minimum QoE and which one is better for both operator and users. If the guaranteed service is in terms of average score during connection holding time then it is acceptable to have a few moments of low QoE and some other high-QoE moments to compensate. On the other hand, if the guaranteed service is in terms of minimum score then network operator needs to make sure during the whole connection, user will perceive at least this minimum value. It needs to be repeated that quality of experience is subjective and, generally, a user is very sensitive to bad quality. This means, a user usually pays more attention on the moment of bad QoE instead of reasonably average the overall quality during the connection. The corresponding mechanisms must take into account this policy in order to set value/threshold that should be used for making decision. In any case, an appropriate SLA needs to be established in advance indicating specification of provided service and responsibility of each party.

As for PSQA implementation and usage, it can be noticed that PSQA is a useful tool to measure user experience in real time; however, it needs to be mentioned that even the output of PSQA (i.e. quality of experience) is independent on application and environment but we can notice that the input of PSQA and its methodology are context-specific. Due to its methodology, a validated RNN will work only with the same application and within similar context in which it has been trained. The RNN trained with video streaming application will not be accurate when using to measure VoIP application. Since the two applications have different characteristics, which normally result in different quality-affecting factors (input of RNN). For example, time-related factors (e.g. delay and jitter) are crucial in VoIP but less important in video streaming because of there are some supports such as buffering before the play-out. As for environment, distribution of loss on wireless network is different from those in wired network and can yield inaccurate results if using with wired technology. Therefore, major inconvenient of this approach is the complexity of PSQA methodology, which needs to be considered carefully for every new context. Nevertheless, once PSQA procedures have been conducted successfully, it can be used easily and automatically.

9.3 Perspective

Resource management in the future will progressively rely on QoE as it is an essential factor of user satisfaction. As network becomes heterogeneous, it will be interesting to investigate on management of radio resources in such environment using QoE metric. Heterogeneity does not concern only network technology but also applications, users, devices, etc. With an emergence of various multimedia applications in next generation network (NGN), various traffic runs currently on the network. Service differentiation will be needed in order to handle all types of applications according to their characteristics and requirements. Different treatments are necessary in order to satisfy user experience and to optimize resource utilization.

Therefore, one prospective topic would concern "*service differentiation*" in NGN. Today, each network application has its needs, a tailored service should be provided by network operator in terms of bandwidth requirement, delay sensitivity, etc. The same argument applies to network users as well. High priority users (who generally pay higher price) should have a privileged access to network resources comparing to medium and low priority users respectively. Resource allocation should be aware of these factors. Two representative applications, namely video and voice over IP, could be considered along with background traffic. Management should be based on quality experienced at end-users in order to be more flexible and more efficient than those based on technical parameters. For instance, new scheduling mechanisms could be proposed for providing appropriate quality to each application and user.

Concerning QoE itself, it would be helpful if we can predict user experience (*QoE Provisioning*). Few works have already begun the investigation. This can be done using learning, mapping, or other modeling strategies. If accurate QoE prediction is available, we can imagine whole network system based on it for resource management. Therefore, another interesting topic would be the study of possibility and feasibility to design such a framework. Many issues need to be investigated: broker or entity to control network resource could be necessary, communications between network entities have to be considered as well as billing and security issues, etc. Moreover, as in heterogeneous environment, heterogeneity also concerns other elements; interoperability issue will become crucial and will have to be studied in order to make everything works together smoothly.

Beside the heterogeneity issues, research directions can also continue on other attractive architectures that have been progressively developed such as *Overlay networks*. Examples are Peer-to-Peer, video delivery network (VDN), or even content delivery network (CDN) in the future. With these network architectures, it will be advantageous to explore how resource management can be enhanced using QoE indicator. As in this document, resource management at network and user sides have been presented; furthermore, end-to-end resource management can also present another research direction. As we can imagine, end-to-end controls and adaptations could be improved greatly with valuable information like user experience.

Glossary

3G : Third Generation
3GPP : 3rd Generation Partnership Project
AAA : Authentication, Authorization and Accounting
AARF : Adaptive Auto Rate fallback
AC : Access Control
ACK : Acknowledgment
AHP : Analytical Hierarchy Process
AM : Acknowledged Mode
AP : Access Point
ARQ : Automatic Repeat Request
BER : Bit Error Rate
BS : Base Station
CCK : Complementary Code Keying
CDF : Cumulative Distribution Function
CDN : Content Delivery Network
CIR : Carrier to Interference Ratio
CMPQM : Color Moving Picture Quality Metric
CN : Core Network
CQI : Channel Quality Indicator
DBPSK : Differential Binary Phase Shift Keying
DQPSK : Differential Quadrature Phase Shift Keying
DSCQS : Double Stimulus Quality Scale
DSIS : Double Stimulus Impairment Scale
DVB-H : Digital Video Broadcasting Ű Handheld
DVB-RC : Digital Video Broadcasting Ű Return Channel
DVB-S : Digital Video Broadcasting Ű Satellite
DVB-T : Digital Video Broadcasting Ű Terrestrial
FEC : Forward Error Correction
FIFO : First In First Out
FLC : Fuzzy Logic Controller
GGSN : Gateway GPRS Support Node
GRA : Grey Relational Analysis

GRC : Grey Relational Coefficient
GSM : Global System for Mobile communications
HSDPA : High Speed Downlink Packet Access
HSPA : High Speed Packet Access
HWN : Heterogeneous Wireless Network
IMPL : Implementation
IST : Information Society Technology
LR : Loss Rate
LTE : Digital Video Broadcasting Ű Terrestrial
MAC : Medium Access Control
MADM : Multi-Attribute Decision Making
MEWS : Multiplicative Exponent Weighting
MIH : Media Independent Handover
MIP : Mobile IP
MLBS : Mean Loss Burst Size
MN : Mobile Node
MNB Measuring Normalizing Block
MOS : Mean Opinion Score
MPQM : EPFL's Moving Picture Quality Metric
MPQM : Moving Picture Quality Metric
MSE : Mean Squared Error
NAK : Negative Acknowledgment
NGN : Next Generation Network
NRG : Normalized Rate Guarantee
NRT : Non Real-time
NS : Network Simulator
NVFM : Normalization Video Fidelity Metric
OSI : Open System Interconnection
PESQ : Perceptual Evaluation of Speech Quality
PF : Proportionally Fair
PoA : Point of Attachment
PSNR : Peak Signal to Noise Ratio
PSQA : Pseudo-Subjective Quality Assessment
PSQM : Perceptual Speech Quality Measure
QoE : Quality of Experience
QoS : Quality of Service
RG : Rate Guarantee
RLC : Radio Link Control Protocol
RM : Resource Management
RNC : Radio Network Controllers
RNN : Random Neural Network

RoHC : Robust Header Compression
RR : Round Robin
RRM : Radio Resource Management
RSS : Received signal strength
RT : Real-time
SAW : Simple Additive Weighting
SCACJ : Stimulus Comparison Adjectival Categorical Judgment
SCS-PD : Single common Service with probabilistic demands
SDSCE : Simultaneous Double Stimulus for Continuous Evaluation
SER : Symbol Error Rate
SGSN : Serving GPRS Support Node
SINR : Signal to Interference plus Noise Ration
SIP : Session Initiation Protocol
SIR : Signal to Interference Ratio
SLA : Service Level Agreement
SLP : Stochastic Linear Programming
SNR : Signal to Noise Ratio
SP : Stochastic Programming
SS : Single Stimulus
SSCQE : Single Stimulus Continuous Quality Evaluation
SVC : Scalable Video Coding
TTI : Transmission Time Interval
UE : User Equipment
UM : Unacknowledged Mode
UMTS : Universal Mobile Telecommunications System
UTRAN : UMTS Terrestrial Radio Access Network
VBR : Variable Bit Rate
VDN : Video Delivery Network
VoIP : Voice over IP
VQM : ITS' Video Quality Metric
WiMAX : Worldwide Interoperability for Microwave Access
WLAN : Wireless Local Network
WMAN : Wireless Metropolitan Network
WMN : Wireless Multimedia Network
WPAN : Wireless Personal Network

Bibliography

[1] Kandaraj Piamrat, Adlen Ksentini, Jean-Marie Bonnin, and César Viho. Radio resource management in emerging heterogeneous wireless networks. *Computer Communications*, In Press, Uncorrected Proof:–, 2010. 9, 39

[2] K. Piamrat, K. Singh, A. Ksentini, C. Viho, and J.-M. Bonnin. QoE-aware scheduling for video-streaming in High Speed Downlink Packet Access. In *IEEE Wireless Communications and Networking Conference, 2010. WCNC 2010.*, april 2010. 9, 25, 39

[3] Kandaraj Piamrat, Adlen Ksentini, Jean-Marie Bonnin, and Cesar Viho. Q-DRAM: QoE-Based Dynamic Rate Adaptation Mechanism for Multicast in Wireless Networks. In *IEEE Global Telecommunications Conference, 2009. GLOBECOM 2009.*, pages 1 –6, 30 Nov. - 4 Dec. 2009 2009. 9, 22, 39

[4] K. Piamrat, A. Ksentini, J.-M. Bonnin, and C. Viho. Rate adaptation mechanism for multimedia multicasting in wireless networks. In *Sixth International Conference on Broadband Communications, Networks, and Systems, 2009. BROADNETS 2009.*, pages 1 –7, sept. 2009. 9, 39

[5] K. Piamrat, C. Viho, J.-M. Bonnin, and A. Ksentini. Quality of Experience Measurements for Video Streaming over Wireless Networks. In *Sixth International Conference on Information Technology: New Generations, 2009. ITNG '09.*, pages 1184 –1189, april 2009. 9, 39

[6] K. Piamrat, A. Ksentini, C. Viho, and J.-M. Bonnin. QoE-Aware Admission Control for Multimedia Applications in IEEE 802.11 Wireless Networks. *IEEE 68th Vehicular Technology Conference, 2008. VTC 2008-Fall.*, pages 1–5, Sept. 2008. 10, 21, 40

[7] K. Piamrat, A. Ksentini, C. Viho, and J.-M. Bonnin. QoE-based network selection for multimedia users in IEEE 802.11 wireless networks. *33rd IEEE Conference on Local Computer Networks, 2008. LCN 2008.*, pages 388–394, Oct. 2008. 10, 26, 40

[8] Kandaraj Piamrat, César Viho, Adlen Ksentini, and Jean-Marie Bonnin. QoE-aware Network Selection in Wireless Heterogeneous Networks. Research Report RR-7282, INRIA, 05 2010. 10, 28, 40

[9] ITU-International Telecommunication Union. Definition of Quality of Experience (QoE), January 2007. 10, 77

[10] ITU-International Telecommunication Union. ITU-R Recommendation BT.500-11. Methodology for the subjective assessment of the quality of television pictures. 15, 78

[11] A. N. Netravali and B. G. Haskell. Digital Pictures: Representation, Compression, and Standards). In *(2nd Ed)*, 1995. 16, 80

[12] J. Gross, J. Klaue, H. Karl, and A. Wolisz. Cross-layer optimization of OFDM transmission systems for MPEG-4 video streaming. *Computer Communications*, 27(11):1044–1055, 2004. 17, 81

[13] ITU-T Recommendation G.107. The E-model, a computational model for use in transmission planning. 17, 81

[14] ITU-T Recommendation P.861. Objective quality measurement of telephone-band (300-3400 Hz) speech codecs. 17, 81

[15] ITU-T Recommendation P.862. Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. 17, 81

[16] S. Mohamed and G. Rubino. A study of real-time packet video quality using random neural networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(12):1071–1083, Dec 2002. 17, 141

[17] Erol Gelenbe. Learning in the recurrent random neural network. *Neural Computation*, 5(1):154–164, 1993. 17, 83

[18] E. Gilbert. Capacity of a Burst loss Channel. *Bell Systems Technical Journal*, 5(39), September 1960. 19, 86

[19] Cheng-Han Lin, Chih-Heng Ke, Ce-Kuen Shieh, and Naveen K. Chilamkurti. The Packet Loss Effect on MPEG Video Transmission in Wireless Networks. In *AINA '06: Proceedings of the 20th International Conference on Advanced Information Networking and Applications*, pages 565–572, Washington, DC, USA, 2006. IEEE Computer Society. 20

[20] Helmut Hlavacs, Karin A. Hummel, Alexander Albl, Stefan Frandl, Jürgen Kirchler, and Boris Ilic. *Effects of WLAN QoS Degradation on Streamed MPEG4 Video Quality*, chapter Wireless Systems and Mobility in Next Generation Internet, pages 152–165. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, Feburary 2005. 20

[21] The Network Simulator - NS-2. http://www.isi.edu/nsnam/ns/. 20, 85, 87, 110, 123, 160

[22] EURANE-Enhanced UMTS Radio Access Networks for ns-2 . http://www.ti-wmc.nl/eurane/. 20, 143

[23] National Institute of Standards and Technology (NIST), Advanced Network Technologies Division (ANTD). The Network simulator NS-2 NIST add-on: IEEE 802.21 model. In *(based on IEEE P802.21/D03.00)*, January 2007. http://www.antd.nist.gov/seamlessandsecure/. 20, 169

[24] Anas Basalamah, Hiroki Sugimoto, and Takuro Sato. A Rate Adaptive Multicast Protocol for Providing MAC Layer Reliability in WLANs. *IEICE Transactions on Communications*, E89-B(10):2733–2740, 2006. 22, 118

[25] J. Villalon, P. Cuenca, L. Orozco-Barbosa, Y. Seok, and T. Turletti. ARSM: a cross-layer auto rate selection multicast mechanism for multi-rate wireless LANs. *IET Communications*, 1(5):893–902, Oct. 2007. 22, 119

[26] Sungjoon Choi, Nakjung Choi, Yongho Seok, Taekyoung Kwon, and Yanghee Choi. Leader-Based Rate Adaptive Multicasting for Wireless LANs. In *IEEE Global Telecommunications Conference, 2007. GLOBECOM '07*, pages 3656–3660, Nov. 2007. 22, 118

[27] Youngsam Park, Yongho Seok, Nakjung Choi, Yanghee Choi, and J.-M. Bonnin. Rate-adaptive multimedia multicasting over IEEE 802.11 wireless LANs. *3rd IEEE Consumer Communications and Networking Conference, 2006. CCNC 2006.*, 1:178–182, 8-10 Jan. 2006. 23, 119, 126

[28] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda. Performance anomaly of 802.11b. *INFOCOM, Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies*, 2:836–843 vol.2, March-3 April 2003. 23, 116

[29] P. Hosein. Qos Control for WCDMA High Speed Packet Data. *International Workshop on Mobile and Wireless Network Communications*, pages 169–173, September 2002. 25, 140

[30] K. Singh and D. Ros. Normalized rate guarantee scheduler for High Speed Downlink Packet Access. In *IEEE GLOBECOM*, 2007. 26, 140

[31] The Institute of Electrical and Electronics Engineers (IEEE). Media Independent Handover Services IEEE Standard Draft 802.21, 2007. Rev.4.0. 28, 63, 70, 98, 167

[32] ITU-T. X.200 : Information technology - Open Systems Interconnection - Basic Reference Model: The basic model, July 1994. 37

[33] Kandaraj Piamrat, César Viho, Adlen Ksentini, and Jean-Marie Bonnin. Resource Management in Mobile Heterogeneous Networks: State of the Art and Challenges. Research Report RR-6459, INRIA, 2008. 40

[34] Kandaraj Piamrat, César Viho, Adlen Ksentini, and Jean-Marie Bonnin. Rate Adaptation Mechanisms for Multimedia Multicasting in Wireless IEEE 802.11 Networks. Research report, INRIA, 2009. 40

[35] M. Annoni, R. Hancock, T. Paila, E. Scarrone, R. Tonjes, L. Dell'Uomo, D. Wisely, and R. Mort. Radio access networks beyond 3G: a first comparison of architectures. *12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2001*, 2:G–133–G–140 vol.2, Sep/Oct 2001. 52

[36] Anis Zouari. *Contrôle de la mobilité et de la qualité de service entre des réseaux d'accès hétérogènes dans un réseau de coeur tout IP*. PhD thesis, RSM - Dépt. Réseaux, Sécurité et Multimédia (Institut TELECOM ; TELECOM Bretagne), 2009. 52

[37] Leijia Wu and Kumbesan Sandrasegaran. A Survey on Common Radio Resource Management. *2nd International Conference on Wireless Broadband and Ultra Wideband Communications, 2007. AusWireless 2007.*, pages 66–66, 27-30 Aug. 2007. 52

[38] P. Kall and S. W Wallace. *Stochastic Programming*. Wiley, Chicheste, 1994. 57

[39] A.-E.M. Taha, H.S. Hassanein, and H.T. Mouftah. On robust allocation policies in wireless heterogeneous networks. *First International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks, 2004. QSHINE 2004.*, pages 198–205, 18-20 Oct. 2004. 57, 59, 70

[40] D. Niyato and E. Hossain. A Cooperative Game Framework for Bandwidth Allocation in 4G Heterogeneous Wireless Networks. *IEEE International Conference on Communications, 2006. ICC '06.*, 9:4357–4362, June 2006. 58, 59

[41] Xu Yang, J. Bigham, and L. Cuthbert. Resource management for service providers in heterogeneous wireless networks. *IEEE Wireless Communications and Networking Conference, 2005*, 3:1305–1310, 13-17 March 2005. 59, 60, 68, 70, 166

[42] C.-T. Lin and P.-J. Juan. Applying the AHP and GRA to Location Selecting for International Resort Parks. *Journal of Grey System*, 19(4), 2007. 61

[43] Qingyang Song and A. Jamalipour. A network selection mechanism for next generation networks. *IEEE International Conference on Communications, 2005. ICC 2005.*, 2:1418–1422, 16-20 May 2005. 61, 62, 70, 166

[44] T. L. Saaty. Fundamentals of Decision Making and Priority Theory With the Analytic Hierarchy Process. *Analytic Hierarchy Process Series, RWS Publications, U.S.A.*, 6, 2000. 62, 199

[45] M. Li, S. Chen, and D. Xie. A multi-step vertical handoff mechanism for cellular multi-hop networks. In *the 2nd ACM Workshop on Performance Monitoring and Measurement of Heterogeneous Wireless and Wired Networks*, pages 119–12, October 2007. 62

[46] Charilas D., Markaki O., Nikitopoulos D., and Theologou M. Packet-switched network selection with the highest QoS in 4G networks. *Computer Networks, Elsevier*, 52:248–258, 2008. 62

[47] H. Wang, C. Li, and Y. Ji. On-demand data broadcast scheduling based on AHP and GRA methods in wireless broadcast/UMTS integrated systems. *The Journal of China Universities of Posts and Telecommunications*, 14(4):1–6, December 2007. 62

[48] K. Yoon and C. Hwang. Multiple attribute decision making: an introduction. *Sage University Paper series on Quantitative Applications in the Social Sciences (07-104)*, 1995. 62

[49] O. Ormond, P. Perry, and J. Murphy. Network selection decision in wireless heterogeneous networks. *IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications, 2005. PIMRC 2005.*, 4:2680–2684, 11-14 Sept. 2005. 62, 70

[50] H. Hotelling. Edgeworth's Taxation Paradox and the Nature of Supply and Demand Functions. *Journal of Political Economy*, 1932. 63

[51] Xiaoshan Liu, V.O.K. Li, and Ping Zhang. Joint Radio Resource Management through Vertical Handoffs in 4G Networks. *IEEE Global Telecommunications*

Conference, 2006. GLOBECOM '06., pages 1–5, 27 2006-Dec. 1 2006. 63, 64, 68, 69, 70, 166

[52] A.L. Wilson, A. Lenaghan, and R. Malyan. Optimising Wireless Access Network Selection to Maintain QoS in Heterogeneous Wireless Environments. *Wireless Personal Multimedia Communications, 2005*, 18-22 Sep. 2005. 64, 65, 68, 70, 166

[53] P. Chan, R. Sheriff, Y. Hu, P. Conforto, and C. Tocci. Mobility management incorporating fuzzy logic for a heterogeneous IP environment. *IEEE Communications Magazine*, 39:42–51, 2001. 64

[54] L. Giupponi, R. Agusti, J. Perez-Romero, and O. Sallent. Joint radio resource management algorithm for multi-RAT networks. *IEEE Global Telecommunications Conference, 2005. GLOBECOM '05.*, 6:5 pp.–, 28 Nov.-2 Dec. 2005. 64

[55] R. Agusti. Radio resource management in beyond 3G systems. *IEEE Mediterranean Electrotechnical Conference, 2006. MELECON 2006.*, pages 569–574, May 2006. 64

[56] G. Koundourakis, D.I. Axiotis, and M. Theologou. Network-Based Access Selection in Composite Radio Environments. *IEEE Wireless Communications and Networking Conference, 2007. WCNC 2007.*, pages 3877–3883, 11-15 March 2007. 65, 69, 70, 72

[57] G. A. Norris and Marshall H. E. Multiattribute Decision Analysis Method for Evaluating Buildings and Building Systems. National Institute of Standards and Technology (NIST), September 1995. 66

[58] A. Ben Nacef and N. Montavont. A generic end-host mechanism for path selection and flow distribution. In *IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC 2008*, pages 1–5, 2008. 66, 69

[59] E. Stevens-Navarro and V.W.S. Wong. Comparison between Vertical Handoff Decision Algorithms for Heterogeneous Wireless Networks. In *IEEE 63rd Vehicular Technology Conference, 2006. VTC 2006-Spring.*, volume 2, pages 947–951, May 2006. 66

[60] Gustavo Carneiro, Carlos García, Pedro Neves, Zhiqui Chen, Michelle Wetterwald, Manuel Ricardo, Pablo Serrano, Susana Sargento, and Albert Banchs. The DAIDALOS Architecture for QoS over Heterogeneous Wireless Networks. In *Proceedings of the 14th IST Mobile and Wireless Communications Summit*, June 2005. 68, 69, 73

[61] Yen-Wen Chen, I-Hsuan Peng, and Shiu-Tsuen Guan. Dynamic Bandwidth Management for Handoffs with RSVP in 802.16/WLAN Environment. *21st International Conference on Advanced Information Networking and Applications Workshops, 2007, AINAW '07.*, 2:243–248, 21-23 May 2007. 68, 70

[62] G.J. Hoekstra, O. Østerbø, R. Schwendener, J. Schneider, F.J.M. Panken, and J. van Bemmel. Quality of Service Solution for Open Wireless Access Networks. *European Signal Processing Conference, EUSIPCO 2005*, 4-8 Sep. 2005. 69

[63] Xia Gao, Gang Wu, and Toshio Miki. QoS framework for mobile heterogeneous networks. *IEEE International Conference on Communications, 2003. ICC '03.*, 2:933–937 vol.2, May 2003. 69

[64] C. Perkins. IP Mobility Support. *IETF RFC 2002*, 1996. 69

[65] N. Banerjee, W. Wu, K. Basu, and S.K. Das. Analysis of SIP-based mobility management in 4G wireless networks. *Computer Communications*, 27:697–707, 2004. 69

[66] Y. Yeea, K. Choongb, A.L. Lowa, and S. Tana. SIP-based proactive and adaptive mobility management framework for heterogeneous networks. *Computer Communications*, 31:771–792, 2008. 69

[67] I.F. Akyildiz, S. Mohanty, and Jiang Xie. A ubiquitous mobile communication architecture for next-generation heterogeneous wireless systems. *IEEE Communications Magazine*, 43(6):S29–S36, June 2005. 69, 72

[68] J.-M. Bonnin, I. Lassoued, and Z.B. Hamouda. Automatic multi-interface management through profile handling. *Mobile Networks and Applications, Kluwer Academic Publishers*, 14:4–17, 2009. 69

[69] I. Lassoued, J.-M. Bonnin, and A. Belghith. Towards an Architecture for Mobility Management and Resource Control. In *IEEE Wireless Communications and Networking Conference, 2008. WCNC 2008*, pages 2846–2851, 2008. 69, 98

[70] P. Magnusson, J. Lundsjo, J. Sachs, and P. Wallentin. Radio resource management distribution in a beyond 3G multi-radio access architecture. *IEEE Global Telecommunications Conference, 2004. GLOBECOM '04*, 6:3472–3477, 2004. 72

[71] H. Schwarz, D. Marpe, and T. Wiegand. Overview of the Scalable Video Coding Extension of the H.264/AVC Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(9):1103 –1120, sept. 2007. 72

[72] Dominique Pichon, Jean-Marie Bonnin, and Pierrick Seite. Context-aware delivery of video content to mobile users. In ACM, editor, *International Conference On Mobile Technology, Applications, And Systems archive*. FT - France Telecom, Division R&D, Rennes (France Télécom), RSM - Dépt. Réseaux, Sécurité et Multimédia (Institut Télécom-Télécom Bretagne), 2009. 72

[73] Dominique Pichon, Guillaume Rieublandou, Jean-Marie Bonnin, and Pierrick Seite. Coordinated service and mobility management to maximize user qoe and optimize network resource use. In *PIMRC'09 : IEEE international symposium on personal, indoor and mobile radio communications*. FT - France Telecom, Division R&D, Rennes (France Télécom), RSM - Dépt. Réseaux, Sécurité et Multimédia (Institut Télécom-Télécom Bretagne), 2009. 72

[74] A. Kassler, A. Schorr, L. Chen, C. Niedermeier, C. Meyer, M. Helbing, and M. Talanda. Multimedia communication in policy based heterogeneous wireless networks. *IEEE 59th Vehicular Technology Conference, 2004. VTC 2004-Spring. 2004*, 5:2987–2992, May 2004. 72

[75] K. D. Singh. *Improving Quality of Service and Resource Utilisation for Multimedia Streaming over Third Generation Mobile Networks*. PhD thesis, University of Rennes 1, 2007. 72

[76] Pablo Rodriguez-Bocca. *Quality-centric design of Peer-to-Peer systems for live-video broadcasting*. PhD thesis, Ecole Doctorale: Matisse, IFSIC, University of Rennes 1, February 2008. 79, 199

[77] S. Winkler and P. Mohandas. The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics. *IEEE Transactions on Broadcasting*, 54(3):660 –668, sept. 2008. 81

[78] Gerardo Rubino. Quantifying the Quality of Audio and Video Transmissions over the Internet: the PSQA Approach. *Design and Operations of Communication Networks: A Review of Wired and Wireless Modelling and Management Challenges, Imperial College Press*, 2005. 81

[79] Jeff Heaton. *Introduction to Neural Networks with Java, Second Edition*. Heaton Research, Inc., 2004. 82

[80] IEEE 802.11 WG. IEEE 802.11, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. *IEEE Std 802.11-2007 (Revision of IEEE Std 802.11-1999)*, pages C1–1184, 12 June 2007. 85, 98, 105, 117

[81] Alessio Botta, Alberto Dainotti, and Antonio Pescapè. Multi-protocol and multi-platform traffic generation and measurement. *INFOCOM 2007 DEMO Session, Anchorage (Alaska, USA)*, May, 2007. 86

[82] Jarmo Prokkola, Mikko Hanski, Marko Jurvansuu, and Milla Immonen. HSDPA Performance in Live Networks. In *IEEE International Conference on Communications*, June 24-28 2007. 86

[83] Marco Fiore. NS-2.29 Wireless Update Patch. http://www.tlc-networks.polito.it/fiore/. 87, 160

[84] Gerardo Rubino. The PSQA project. http://www.irisa.fr/armor/lesmembres/Rubino/myPages/psqa.html. 87

[85] Gerardo Rubino, Martín Varela, and Jean-Marie Bonnin. Controlling Multimedia QoS in the Future Home Network Using the PSQA Metric. *Computer Journal*, 49(2):137–155, 2006. 95

[86] H. Cancela, P. Rodriguez-Bocca, and G. Rubino. Perceptual Quality in P2P Multi-Source Video Streaming Policies. *IEEE Global Telecommunications Conference, 2007. GLOBECOM '07.*, pages 2780–2785, Nov. 2007. 95, 108, 120

[87] A.P. da Silva, P. Rodriguez-Bocca, and G. Rubino. Optimal Quality-of-Experience Design for a P2P Multi-Source Video Streaming. In *IEEE International Conference on Communications, 2008. ICC '08.*, pages 22 –26, may 2008. 98

[88] IEEE 802.11 WG. Unapproved Draft Amendment Standard for Information Technology– Telecommunications and Information Exchange Between Systems–LAN/MAN Specific Requirements– Part 11 Wireless Medium Access Control (MAC) and Physical Layer (PHY) specifications: Medium Access Control (MAC) Quality of Service (QoS) Enhancements (Replaced by approved draft 802.11e/D13.0). *IEEE Std P802.11e/D13.0*, pages –, 2005. 106

[89] Shiann-Tsong Sheu and Tzu-Fang Sheu. A bandwidth allocation/sharing/extension protocol for multimedia over IEEE 802.11 ad hoc wireless LANs. *IEEE Journal on Selected Areas in Communications*, 19(10):2065–2080, Oct 2001. 106

[90] G. Bianchi. Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE Journal on Selected Areas in Communications*, 18(3):535–547, March 2000. 106

[91] D. Pong and T. Moors. Call admission control for IEEE 802.11 contention access mechanism. *IEEE Global Telecommunications Conference, 2003. GLOBECOM '03.*, 1:174–178, 1-5 Dec. 2003. 106

[92] Hongqiang Zhai, Xiang Chen, and Yuguang Fang. A call admission and rate control scheme for multimedia support over IEEE 802.11 wireless LANs. *Wireless Networks*, 12(4):451–463, 2006. 106

[93] P. Rodriguez-Bocca. *Quality-centric design of Peer-to-Peer systems for live-video broadcasting*. PhD thesis, University Rennes 1, 2008. 108

[94] IEEE 802.11 WG. Supplement to IEEE Standard for Information technology - Telecommunication and information exchange between systems - Local and metropolitan area networks - Specific requirements. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Higher-Speed Physical Layer Extension in the 2.4 GHz Band. IEEE std 802.11b-1999 (R2003). 110, 117

[95] TKN - Telecommunication and Networks Group. http://www.tkn.tu-berlin.de/. 110

[96] J. Kuri and S.K. Kasera. Reliable multicast in multi-access wireless LANs. *Proceedings. IEEE INFOCOM'99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies.*, 2:760–767 vol.2, Mar 1999. 116

[97] Min-Te Sun, Lifei Huang, Anish Arora, and Ten-Hwang Lai. Reliable MAC Layer Multicast in IEEE 802.11 Wireless Networks. In *ICPP '02: Proceedings of the 2002 International Conference on Parallel Processing (ICPP'02)*, page 527, Washington, DC, USA, 2002. IEEE Computer Society. 116

[98] IEEE 802.11 WG. IEEE standard for information technology- telecommunications and information exchange between systems- local and metropolitan area networks- specific requirements Part II: wireless LAN medium access control (MAC) and physical layer (PHY) specifications, 2003. 117

[99] A. Kamerman and L. Monteban. WaveLAN-II: A high-performance wireless LAN for the unlicensed band. *Bell Labs technial journal*, 2(3), 1997. 117

[100] Mathieu Lacage, Mohammad Hossein Manshaei, and Thierry Turletti. IEEE 802.11 rate adaptation: a practical approach. In *MSWiM: Proceedings of the 7th ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems*, pages 126–134, New York, NY, USA, 2004. ACM. 117

[101] Gavin Holland, Nitin Vaidya, and Paramvir Bahl. A rate-adaptive MAC protocol for multi-Hop wireless networks. In *MobiCom: Proceedings of the 7th annual international conference on Mobile computing and networking*, pages 236–251, New York, NY, USA, 2001. ACM. 118

[102] IEEE Computer Society. IEEE 802.11k, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 1: Radio Resource Measurement of Wireless LANs. *IEEE Std 802.11k-2008 (Amendment to IEEE Std 802.11-2007)*, pages c1–222, 12 June 2008. 120, 156

[103] O. Nemethova, M. Ries, M. Rupp, and E. Siffel. Quality Assessment for H.264 Coded Low-Rate and Low-Resolution Video Sequences. In *Conference on Internet and Information Technologies (CIIT)*, pages 136–140, 2004. 128

[104] H. Holma and A. Toskala. *WCDMA for UMTS: Radio Access for Third Generation Mobile Communications, 3rd ed.* Wiley, 2004. 138, 139

[105] F. Agboma and A. Liotta. QoE-aware QoS management. In *The 6th International Conference on Advances in Mobile Computing and Multimedia, MoMM'2008*, pages 111–116, November 24-26 2008. 139

[106] V. Menkovski, A. Oredope, A. Liotta, and A. Cuadra Sanchez. Predicting quality of experience in multimedia streaming. In *7th International Conference on Advances in Mobile Computing and Multimedia (MoMM2009)*, December 14-16 2009. 139

[107] A. Jalali, R. Padovani, and R. Pankaj. Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system. In *IEEE VTC 2000 Spring*, volume 3, pages 1663–166, 2000. 140

[108] F. Kelly. Charging and Rate Control for Elastic Traffic. *European Transactions on Telecommunications*, 8:33–37, 1997. 140

[109] T. E. Kolding. QoS-Aware Proportional Fair Packet Scheduling with Required Activity Detection. In *IEEE VTC 2006-Fall*, September 2006. 140

[110] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, , and R. Vijayakumar. Providing quality of service over a shared wireless link. *IEEE Communications Magazine*, 39(2):150–154, February 2001. 140

[111] Raj Jain, Dah-Ming Chiu, and W. Hawe. A Quantitative Measure Of Fairness And Discrimination For Resource Allocation In Shared Computer Systems. *CoRR - Computing Research Repository*, cs.NI/9809099, 1998. 163

[112] Birdstep Technology. Secgo Mobile IP. http://www.secgo.com/. 169

List of Figures

List of Tables

Résumé

Les applications multimédias pour terminaux mobiles connaissent un succès grandissant. Cela oblige à développer de nouvelles méthodes plus efficaces de gestion des ressources des réseaux sans-fil du fait de leurs caractéristiques particulières : bande-passante limitée, état radio variable, interférences plus importantes, etc. Par ailleurs, les méthodes classiques de la gestion de ressources basées sur des paramètres techniques (perte/retard de paquets, gigue, etc.) ne parviennent pas à donner des évaluations précises de la qualité telle que perçue (encore appelée Qualité d'Expérience ou QdE) par l'utilisateur de ces applications. Cette thèse s'appuie sur une technique hybride nommée PSQA (Pseudo-Subjective Quality Assessment) d'évaluation pseudo-subjective en temps réel de la QdE pour proposer de nouvelles méthodes de gestion de ressources dans les réseaux multimédias sans-fil. Que ce soit du côté de l'opérateur réseau ou du côté de l'utilisateur, nous avons proposé des méthodes de contrôle d'accès et d'ordonnancement ainsi que des méthodes de sélection de réseaux d'accès dans le contexte des réseaux sans-fil hétérogènes utilisant différentes technologies (IEEE 802.11, UMTS, etc.). Les résultats obtenus encouragent l'utilisation du concept de QdE et ouvre la voie à un nouveau paradigme dans la gestion des ressources dans les réseaux multimédias sans-fil.

Mot clé: Gestion de ressources, Réseaux sans-fil, Qualité d'Expérience, Application multimédia, Réseaux hétérogènes

Abstract

Wireless multimedia networking is gaining tremendous success nowadays. Due to their characteristics (limited bandwidth, variable radio conditions, greater interference, etc.), the need of more efficient management has become crucial. Meanwhile, traditional ways of managing network, using information from monitoring technical parameters (loss, delays, jitter, etc.), fail to give accurate evaluations of user experience or Quality of Experience (QoE). In this thesis, new methods based on QoE indicator have been proposed to solve these problems. The propositions are admission control, rate adaptation, and packet scheduling regarding network operator as well as network selection regarding user side. The real-time measurement of QoE is accomplished with PSQA (Pseudo-Subjective Quality Assessment) tool. The simulations have been conducted using different wireless technologies both in homogeneous and heterogeneous environment. The obtained results encourage the use of QoE concept in further research, which could pave the road to a new paradigm of resource management.

Keyword: Resource Management, Wireless Networks, Quality of Experience, Multimedia Applications, Heterogeneous Networks